

(19) World Intellectual Property Organization  
International Bureau(43) International Publication Date  
14 February 2002 (14.02.2002)

PCT

(10) International Publication Number  
WO 02/12467 A2(51) International Patent Classification<sup>7</sup>: C12N 9/00

(21) International Application Number: PCT/US01/24382

(22) International Filing Date: 3 August 2001 (03.08.2001)

(25) Filing Language: English

(26) Publication Language: English

## (30) Priority Data:

60/223,055	4 August 2000 (04.08.2000)	US
60/224,728	11 August 2000 (11.08.2000)	US
60/226,440	18 August 2000 (18.08.2000)	US
60/228,067	24 August 2000 (24.08.2000)	US
60/230,063	31 August 2000 (31.08.2000)	US
60/232,244	13 September 2000 (13.09.2000)	US
60/234,269	20 September 2000 (20.09.2000)	US

CA 94025 (US). RAMKUMAR, Jayalaxmi [IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US). RING, Huijun, Z. [US/US]; 625 Orange Avenue, Los Altos, CA 94022 (US). SANJANWALA, Madhu, S. [US/US]; 210 Sylvia Court, Los Altos, CA 94024 (US). TANG, Y., Tom [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). THANGAVELU, Kavitha [IN/US]; 1950 Montecito Avenue 23, Mountain View, CA 94043 (US). THORNTON, Michael [US/US]; 9 Medway Road, Woodside, CA 94062 (US). TRIBOULEY, Catherine, M. [FR/US]; 1121 Tennessee Street, #5, San Francisco, CA 94107 (US). WALIA, Narinder, K. [US/US]; 890 Davis Street #205, San Leandro, CA 94577 (US). WARREN, Bridget, A. [US/US]; 10130 Parwood Drive #2, Cupertino, CA 95014 (US). YANG, Junming [CN/US]; 1739 Bank Lane, San Jose, CA 95129 (US). YAO, Monique, G. [US/US]; 111 Frederick Court, Mountain View, CA 94043 (US). YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US).

(71) Applicant (for all designated States except US): INCYTE GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

## (72) Inventors; and

(75) Inventors/Applicants (for US only): BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). BRUNS, Christopher, M. [US/US]; 575 S. Rengstorff Avenue #126, Mountain View, CA 94040 (US). DAS, Debopriya [IN/US]; 1179 Bonita Avenue, Apt. 3, Mountain View, CA 94040 (US). DELEGEANE, Angelo, M. [US/US]; 594 Angus Drive, Milpitas, CA 95035 (US). DING, Li [CN/US]; 3353 Alma Street, #146, Palo Alto, CA 94306 (US). ELLIOT, Vicki, S. [US/US]; 3770 Polton Place Way, San Jose, CA 95121 (US). GANDHI, Ameena, R. [US/US]; 837 Roble Avenue, #1, Menlo Park, CA 94025 (US). GRIFFIN, Jennifer, A. [US/US]; 33691 Mello Way #11, Fremont, CA 94555 (US). HAFALIA, April, J., A. [US/US]; 2227 Calle de Primavera, Santa Clara, CA 95054 (US). KHAN, Farrah, A. [IN/US]; 3617 Central Road #102, Glenview, Illinois 60025 (US). LAL, Preeti [IN/US]; P.O. Box 5142, Santa Clara, CA 95056 (US). LEE, Sally [US/US]; 825 East Evelyn, #425, Sunnyvale, CA 94086 (US). LU, Dyung, Aina, M. [US/US]; 233 Coy Drive, San Jose, CA 95123 (US). LU, Yan [CN/US]; 3885 Corrina Way, Palo Alto, CA 94303 (US). PATTERSON, Chandra [US/US]; 490 Sherwood Way #1, Menlo Park,

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

## Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: DRUG METABOLIZING ENZYMES

(57) Abstract: The invention provides human drug metabolizing enzymes (DME) and polynucleotides which identify and encode DME. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of DME.

## DRUG METABOLIZING ENZYMES

### TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of drug metabolizing enzymes  
5 and to the use of these sequences in the diagnosis, treatment, and prevention of  
autoimmune/inflammatory, cell proliferative, developmental, endocrine, eye, metabolic, and  
gastrointestinal disorders, including liver disorders, and in the assessment of the effects of exogenous  
compounds on the expression of nucleic acid and amino acid sequences of drug metabolizing enzymes.

### 10 BACKGROUND OF THE INVENTION

The metabolism of a drug and its movement through the body (pharmacokinetics) are  
important in determining its effects, toxicity, and interactions with other drugs. The three processes  
governing pharmacokinetics are the absorption of the drug, distribution to various tissues, and  
elimination of drug metabolites. These processes are intimately coupled to drug metabolism, since a  
15 variety of metabolic modifications alter most of the physicochemical and pharmacological properties of  
drugs, including solubility, binding to receptors, and excretion rates. The metabolic pathways which  
modify drugs also accept a variety of naturally occurring substrates such as steroids, fatty acids,  
prostaglandins, leukotrienes, and vitamins. The enzymes in these pathways are therefore important  
sites of biochemical and pharmacological interaction between natural compounds, drugs, carcinogens,  
20 mutagens, and xenobiotics.

It has long been appreciated that inherited differences in drug metabolism lead to drastically  
different levels of drug efficacy and toxicity among individuals. For drugs with narrow therapeutic  
indices, or drugs which require bioactivation (such as codeine), these polymorphisms can be critical.  
Moreover, promising new drugs are frequently eliminated in clinical trials based on toxicities which  
25 may only affect a segment of the patient group. Advances in pharmacogenomics research, of which  
drug metabolizing enzymes constitute an important part, are promising to expand the tools and  
information that can be brought to bear on questions of drug efficacy and toxicity (See Evans, W. E.  
and R. V. Relling (1999) *Science* 286:487-491).

Drug metabolic reactions are categorized as Phase I, which functionalize the drug molecule  
30 and prepare it for further metabolism, and Phase II, which are conjugative. In general, Phase I  
reaction products are partially or fully inactive, and Phase II reaction products are the chief excreted  
species. However, Phase I reaction products are sometimes more active than the original  
administered drugs; this metabolic activation principle is exploited by pro-drugs (e.g. L-dopa).

Additionally, some nontoxic compounds (e.g. aflatoxin, benzo[a]pyrene) are metabolized to toxic intermediates through these pathways. Phase I reactions are usually rate-limiting in drug metabolism. Prior exposure to the compound, or other compounds, can induce the expression of Phase I enzymes however, and thereby increase substrate flux through the metabolic pathways. (See Klaassen, C. D.,

- 5 Amdur, M. O. and J. Doull (1996) Casarett and Doull's Toxicology: The Basic Science of Poisons, McGraw-Hill, New York, NY, pp. 113-186; B. G. Katzung (1995) Basic and Clinical Pharmacology, Appleton and Lange, Norwalk, CT, pp. 48-59; G. G. Gibson and P. Skett (1994) Introduction to Drug Metabolism, Blackie Academic and Professional, London.)

- Drug metabolizing enzymes (DMEs) have broad substrate specificities. This can be  
10 contrasted to the immune system, where a large and diverse population of antibodies are highly specific for their antigens. The ability of DMEs to metabolize a wide variety of molecules creates the potential for drug interactions at the level of metabolism. For example, the induction of a DME by one compound may affect the metabolism of another compound by the enzyme.

- DMEs have been classified according to the type of reaction they catalyze and the cofactors  
15 involved. The major classes of Phase I enzymes include, but are not limited to, cytochrome P450 and flavin-containing monooxygenase. Other enzyme classes involved in Phase I-type catalytic cycles and reactions include, but are not limited to, NADPH cytochrome P450 reductase (CPR), the microsomal cytochrome b5/NADH cytochrome b5 reductase system, the ferredoxin/ferredoxin reductase redox pair, aldo/keto reductases, and alcohol dehydrogenases. The major classes of Phase II enzymes  
20 include, but are not limited to, UDP glucuronyltransferase, sulfotransferase, glutathione S-transferase, N-acyltransferase, and N-acetyl transferase.

#### Cytochrome P450 and P450 catalytic cycle-associated enzymes

- Members of the cytochrome P450 superfamily of enzymes catalyze the oxidative metabolism of a variety of substrates, including natural compounds such as steroids, fatty acids, prostaglandins,  
25 leukotrienes, and vitamins, as well as drugs, carcinogens, mutagens, and xenobiotics. Cytochromes P450, also known as P450 heme-thiolate proteins, usually act as terminal oxidases in multi-component electron transfer chains, called P450-containing monooxygenase systems. Specific reactions catalyzed include hydroxylation, epoxidation, N-oxidation, sulfoxidation, N-, S-, and O-dealkylations, desulfation, deamination, and reduction of azo, nitro, and N-oxide groups. These reactions are  
30 involved in steroidogenesis of glucocorticoids, cortisols, estrogens, and androgens in animals; insecticide resistance in insects; herbicide resistance and flower coloring in plants; and environmental bioremediation by microorganisms. Cytochrome P450 actions on drugs, carcinogens, mutagens, and xenobiotics can result in detoxification or in conversion of the substance to a more toxic product.

Cytochromes P450 are abundant in the liver, but also occur in other tissues; the enzymes are located in microsomes. (See ExPASY ENZYME EC 1.14.14.1; Prosite PDOC00081 Cytochrome P450 cysteine heme-iron ligand signature; PRINTS EP450I E-Class P450 Group I signature; Graham-Lorence, S. and Peterson, J.A. (1996) FASEB J. 10:206-214.)

5 Four hundred cytochromes P450 have been identified in diverse organisms including bacteria, fungi, plants, and animals (Graham-Lorence, supra). The B-class is found in prokaryotes and fungi, while the E-class is found in bacteria, plants, insects, vertebrates, and mammals. Five subclasses or groups are found within the larger family of E-class cytochromes P450 (PRINTS EP450I E-Class P450 Group I signature).

10 All cytochromes P450 use a heme cofactor and share structural attributes. Most cytochromes P450 are 400 to 530 amino acids in length. The secondary structure of the enzyme is about 70% alpha-helical and about 22% beta-sheet. The region around the heme-binding site in the C-terminal part of the protein is conserved among cytochromes P450. A ten amino acid signature sequence in this heme-iron ligand region has been identified which includes a conserved cysteine  
15 involved in binding the heme iron in the fifth coordination site. In eukaryotic cytochromes P450, a membrane-spanning region is usually found in the first 15-20 amino acids of the protein, generally consisting of approximately 15 hydrophobic residues followed by a positively charged residue. (See Prosite PDOC00081, supra; Graham-Lorence, supra.)

Cytochrome P450 enzymes are involved in cell proliferation and development. The enzymes  
20 have roles in chemical mutagenesis and carcinogenesis by metabolizing chemicals to reactive intermediates that form adducts with DNA (Nebert, D.W. and Gonzalez, F.J. (1987) Ann. Rev. Biochem. 56:945-993). These adducts can cause nucleotide changes and DNA rearrangements that lead to oncogenesis. Cytochrome P450 expression in liver and other tissues is induced by xenobiotics such as polycyclic aromatic hydrocarbons, peroxisomal proliferators, phenobarbital, and the  
25 glucocorticoid dexamethasone (Dogra, S.C. et al. (1998) Clin. Exp. Pharmacol. Physiol. 25:1-9). A cytochrome P450 protein may participate in eye development as mutations in the P450 gene CYP1B1 cause primary congenital glaucoma (Online Mendelian Inheritance in Man (OMIM) \*601771 Cytochrome P450, subfamily I (dioxin-inducible), polypeptide 1; CYP1B1).

Cytochromes P450 are associated with inflammation and infection. Hepatic cytochrome P450  
30 activities are profoundly affected by various infections and inflammatory stimuli, some of which are suppressed and some induced (Morgan, E.T. (1997) Drug Metab. Rev. 29:1129-1188). Effects observed *in vivo* can be mimicked by proinflammatory cytokines and interferons. Autoantibodies to two cytochrome P450 proteins were found in patients with autoimmune polyendocrinopathy-

candidiasis-ectodermal dystrophy (APECED), a polyglandular autoimmune syndrome (OMIM \*240300 Autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy).

Mutations in cytochromes P450 have been linked to metabolic disorders, including congenital adrenal hyperplasia, the most common adrenal disorder of infancy and childhood; pseudovitamin D-deficiency rickets; cerebrotendinous xanthomatosis, a lipid storage disease characterized by progressive neurologic dysfunction, premature atherosclerosis, and cataracts; and an inherited resistance to the anticoagulant drugs coumarin and warfarin (Isselbacher, K.J. et al. (1994) Harrison's Principles of Internal Medicine, McGraw-Hill, Inc. New York, NY, pp. 1968-1970; Takeyama, K. et al. (1997) *Science* 277:1827-1830; Kitanaka, S. et al. (1998) *N. Engl. J. Med.* 338:653-661; OMIM \*213700 Cerebrotendinous xanthomatosis; and OMIM #122700 Coumarin resistance). Extremely high levels of expression of the cytochrome P450 protein aromatase were found in a fibrolamellar hepatocellular carcinoma from a boy with severe gynecomastia (feminization) (Agarwal, V.R. (1998) *J. Clin. Endocrinol. Metab.* 83:1797-1800).

The cytochrome P450 catalytic cycle is completed through reduction of cytochrome P450 by NADPH cytochrome P450 reductase (CPR). Another microsomal electron transport system consisting of cytochrome b5 and NADPH cytochrome b5 reductase has been widely viewed as a minor contributor of electrons to the cytochrome P450 catalytic cycle. However, a recent report by Lamb, D. C. et al. (1999; *FEBS Lett.* 462:283-8) identifies a Candida albicans cytochrome P450 (CYP51) which can be efficiently reduced and supported by the microsomal cytochrome b5/NADPH cytochrome b5 reductase system. Therefore, there are likely many cytochromes P450 which are supported by this alternative electron donor system.

Cytochrome b5 reductase is also responsible for the reduction of oxidized hemoglobin (methemoglobin, or ferrihemoglobin, which is unable to carry oxygen) to the active hemoglobin (ferrohemoglobin) in red blood cells. Methemoglobinemia results when there is a high level of oxidant drugs or an abnormal hemoglobin (hemoglobin M) which is not efficiently reduced. Methemoglobinemia can also result from a hereditary deficiency in red cell cytochrome b5 reductase (Reviewed in Mansour, A. and Lurie, A. A. (1993) *Am. J. Hematol.* 42:7-12).

Members of the cytochrome P450 family are also closely associated with vitamin D synthesis and catabolism. Vitamin D exists as two biologically equivalent prohormones, ergocalciferol (vitamin D<sub>2</sub>), produced in plant tissues, and cholecalciferol (vitamin D<sub>3</sub>), produced in animal tissues. The latter form, cholecalciferol, is formed upon the exposure of 7-dehydrocholesterol to near ultraviolet light (i.e., 290-310 nm), normally resulting from even minimal periods of skin exposure to sunlight (reviewed in Miller, W.L. and Portale, A.A. (2000) *Trends Endocrinol. Metab.* 11:315-319).

Both prohormone forms are further metabolized in the liver to 25-hydroxyvitamin D (25(OH)D) by the enzyme 25-hydroxylase. 25(OH)D is the most abundant precursor form of vitamin D which must be further metabolized in the kidney to the active form, 1 $\alpha$ ,25-dihydroxyvitamin D (1 $\alpha$ ,25(OH)<sub>2</sub>D), by the enzyme 25-hydroxyvitamin D 1 $\alpha$ -hydroxylase (1 $\alpha$ -hydroxylase). Regulation of

5 1 $\alpha$ ,25(OH)<sub>2</sub>D production is primarily at this final step in the synthetic pathway. The activity of 1 $\alpha$ -hydroxylase depends upon several physiological factors including the circulating level of the enzyme product (1 $\alpha$ ,25(OH)<sub>2</sub>D) and the levels of parathyroid hormone (PTH), calcitonin, insulin, calcium, phosphorus, growth hormone, and prolactin. Furthermore, extrarenal 1 $\alpha$ -hydroxylase activity has been reported, suggesting that tissue-specific, local regulation of 1 $\alpha$ ,25(OH)<sub>2</sub>D production may

10 also be biologically important. The catalysis of 1 $\alpha$ ,25(OH)<sub>2</sub>D to 24,25-dihydroxyvitamin D (24,25(OH)<sub>2</sub>D), involving the enzyme 25-hydroxyvitamin D 24-hydroxylase (24-hydroxylase), also occurs in the kidney. 24-hydroxylase can also use 25(OH)D as a substrate (Shinki, T. *et al.* (1997) *Proc. Natl. Acad. Sci. U.S.A.* 94:12920-12925; Miller, W.L. and Portale, A.A. *supra*; and references within).

15 Vitamin D 25-hydroxylase, 1 $\alpha$ -hydroxylase, and 24-hydroxylase are all NADPH-dependent, type I (mitochondrial) cytochrome P450 enzymes that show a high degree of homology with other members of the family. Vitamin D 25-hydroxylase also shows a broad substrate specificity and may also perform 26-hydroxylation of bile acid intermediates and 25, 26, and 27-hydroxylation of cholesterol (Dilworth, F.J. *et al.* (1995) *J. Biol. Chem.* 270:16766-16774; Miller, W.L. and Portale, A.A. *supra*; and references within).

20

The active form of vitamin D (1 $\alpha$ ,25(OH)<sub>2</sub>D) is involved in calcium and phosphate homeostasis and promotes the differentiation of myeloid and skin cells. Vitamin D deficiency resulting from deficiencies in the enzymes involved in vitamin D metabolism (e.g., 1 $\alpha$ -hydroxylase) causes hypocalcemia, hypophosphatemia, and vitamin D-dependent (sensitive) rickets, a disease characterized

25 by loss of bone density and distinctive clinical features, including bandy or bow leggedness accompanied by a waddling gait. Deficiencies in vitamin D 25-hydroxylase cause cerebrotendinous xanthomatosis, a lipid-storage disease characterized by the deposition of cholesterol and cholestanol in the Achilles' tendons, brain, lungs, and many other tissues. The disease presents with progressive neurologic dysfunction, including postpubescent cerebellar ataxia, atherosclerosis, and cataracts.

30 Vitamin D 25-hydroxylase deficiency does not result in rickets, suggesting the existence of alternative pathways for the synthesis of 25(OH)D (Griffin, J.E. and Zerwekh, J.E. (1983) *J. Clin. Invest.* 72:1190-1199; Gamblin, G.T. *et al.* (1985) *J. Clin. Invest.* 75:954-960; and W.L. and Portale, A.A. *supra*).

Ferredoxin and ferredoxin reductase are electron transport accessory proteins which support at least one human cytochrome P450 species, cytochrome P450c27 encoded by the CYP27 gene (Dilworth, F. J. et al. (1996) Biochem. J. 320:267-71). A *Streptomyces griseus* cytochrome P450, CYP104D1, was heterologously expressed in *E. coli* and found to be reduced by the endogenous ferredoxin and ferredoxin reductase enzymes (Taylor, M. et al. (1999) Biochem. Biophys. Res. Commun. 263:838-42), suggesting that many cytochrome P450 species may be supported by the ferredoxin/ferredoxin reductase pair. Ferredoxin reductase has also been found in a model drug metabolism system to reduce actinomycin D, an antitumor antibiotic, to a reactive free radical species (Flitter, W. D. and Mason, R. P. (1988) Arch. Biochem. Biophys. 267:632-9).

10 Flavin-containing monooxygenase (FMO)

Flavin-containing monooxygenases oxidize the nucleophilic nitrogen, sulfur, and phosphorus heteroatom of an exceptional range of substrates. Like cytochromes P450, FMOs are microsomal and use NADPH and O<sub>2</sub>; there is also a great deal of substrate overlap with cytochromes P450. The tissue distribution of FMOs includes liver, kidney, and lung.

15 There are five different known isoforms of FMO in mammals (FMO1, FMO2, FMO3, FMO4, and FMO5), which are expressed in a tissue-specific manner. The isoforms differ in their substrate specificities and other properties such as inhibition by various compounds and stereospecificity of reaction. FMOs have a 13 amino acid signature sequence, the components of which span the N-terminal two-thirds of the sequences and include the FAD binding region and the FATGY motif which has been found in many N-hydroxylating enzymes (Stehr, M. et al. (1998) Trends Biochem. Sci. 23:56-57; PRINTS FMOXYGENASE Flavin-containing monooxygenase signature).

Specific reactions include oxidation of nucleophilic tertiary amines to N-oxides, secondary amines to hydroxylamines and nitrones, primary amines to hydroxylamines and oximes, and sulfur-containing compounds and phosphines to S- and P-oxides. Hydrazines, iodides, selenides, and boron-containing compounds are also substrates. Although FMOs appear similar to cytochromes P450 in their chemistry, they can generally be distinguished from cytochromes P450 *in vitro* based on, for example, the higher heat lability of FMOs and the nonionic detergent sensitivity of cytochromes P450; however, use of these properties in identification is complicated by further variation among FMO isoforms with respect to thermal stability and detergent sensitivity.

30 FMOs play important roles in the metabolism of several drugs and xenobiotics. FMO (FMO3 in liver) is predominantly responsible for metabolizing (S)-nicotine to (S)-nicotine N-1'-oxide, which is excreted in urine. FMO is also involved in S-oxygenation of cimetidine, an H<sub>2</sub>-antagonist widely used for the treatment of gastric ulcers. Liver-expressed forms of FMO are not under the same regulatory

control as cytochrome P450. In rats, for example, phenobarbital treatment leads to the induction of cytochrome P450, but the repression of FMO1.

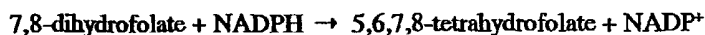
Endogenous substrates of FMO include cysteamine, which is oxidized to the disulfide, cystamine, and trimethylamine (TMA), which is metabolized to trimethylamine N-oxide. TMA smells like rotting fish, and mutations in the FMO3 isoform lead to large amounts of the malodorous free amine being excreted in sweat, urine, and breath. These symptoms have led to the designation fish-odor syndrome (OMIM 602079 Trimethylaminuria).

#### Lysyl oxidase:

Lysyl oxidase (lysine 6-oxidase, LO) is a copper-dependent amine oxidase involved in the formation of connective tissue matrices by crosslinking collagen and elastin. LO is secreted as a N-glycosylated precursor protein of approximately 50 kDa and cleaved to the mature form of the enzyme by a metalloprotease, although the precursor form is also active. The copper atom in LO is involved in the transport of electron to and from oxygen to facilitate the oxidative deamination of lysine residues in these extracellular matrix proteins. While the coordination of copper is essential to LO activity, insufficient dietary intake of copper does not influence the expression of the apoenzyme. However, the absence of the functional LO is linked to the skeletal and vascular tissue disorders that are associated with dietary copper deficiency. LO is also inhibited by a variety of semicarbazides, hydrazines, and amino nitrites, as well as heparin. Beta-aminopropionitrile is a commonly used inhibitor. LO activity is increased in response to ozone, cadmium, and elevated levels of hormones released in response to local tissue trauma, such as transforming growth factor-beta, platelet-derived growth factor, angiotensin II, and fibroblast growth factor. Abnormalities in LO activity have been linked to Menkes syndrome and occipital horn syndrome. Cytosolic forms of the enzyme have been implicated in abnormal cell proliferation (reviewed in Rucker, R.B. et al. (1998) Am. J. Clin. Nutr. 67:996S-1002S and Smith-Mungo, L.I. and Kagan, H.M. (1998) Matrix Biol. 16:387-398).

#### Dihydrofolate reductases

Dihydrofolate reductases (DHFR) are ubiquitous enzymes that catalyze the NADPH-dependent reduction of dihydrofolate to tetrahydrofolate, an essential step in the *de novo* synthesis of glycine and purines as well as the conversion of deoxyuridine monophosphate (dUMP) to deoxythymidine monophosphate (dTMP). The basic reaction is as follows:



The enzymes can be inhibited by a number of dihydrofolate analogs, including trimethoprim and

methotrexate. Since an abundance of TMP is required for DNA synthesis, rapidly dividing cells require the activity of DHFR. The replication of DNA viruses (i.e., herpesvirus) also requires high levels of DHFR activity. As a result, drugs that target DHFR have been used for cancer chemotherapy and to inhibit DNA virus replication. (For similar reasons, thymidylate synthetases are also target enzymes.) Drugs that inhibit DHFR are preferentially cytotoxic for rapidly dividing cells (or DNA virus-infected cells) but have no specificity, resulting in the indiscriminate destruction of dividing cells. Furthermore, cancer cells may become resistant to drugs such as methotrexate as a result of acquired transport defects or the duplication of one or more DHFR genes (Stryer, L. (1988) Biochemistry. W.H Freeman and Co., Inc. New York. pp. 511-5619).

#### 10 Aldo/keto reductases

Aldo/keto reductases are monomeric NADPH-dependent oxidoreductases with broad substrate specificities (Bohren, K. M. et al. (1989) *J. Biol. Chem.* 264:9547-51). These enzymes catalyze the reduction of carbonyl-containing compounds, including carbonyl-containing sugars and aromatic compounds, to the corresponding alcohols. Therefore, a variety of carbonyl-containing drugs and xenobiotics are likely metabolized by enzymes of this class.

One known reaction catalyzed by a family member, aldose reductase, is the reduction of glucose to sorbitol, which is then further metabolized to fructose by sorbitol dehydrogenase. Under normal conditions, the reduction of glucose to sorbitol is a minor pathway. In hyperglycemic states, however, the accumulation of sorbitol is implicated in the development of diabetic complications (OMIM \*103880 Aldo-keto reductase family 1, member B1). Members of this enzyme family are also highly expressed in some liver cancers (Cao, D. et al. (1998) *J. Biol. Chem.* 273:11429-35).

#### Alcohol dehydrogenases

Alcohol dehydrogenases (ADHs) oxidize simple alcohols to the corresponding aldehydes. ADH is a cytosolic enzyme, prefers the cofactor NAD<sup>+</sup>, and also binds zinc ion. Liver contains the highest levels of ADH, with lower levels in kidney, lung, and the gastric mucosa.

Known ADH isoforms are dimeric proteins composed of 40 kDa subunits. There are five known gene loci which encode these subunits (a, b, g, p, c), and some of the loci have characterized allelic variants (b<sub>1</sub>, b<sub>2</sub>, b<sub>3</sub>, g<sub>1</sub>, g<sub>2</sub>). The subunits can form homodimers and heterodimers; the subunit composition determines the specific properties of the active enzyme. The holoenzymes have therefore been categorized as Class I (subunit compositions aa, ab, ag, bg, gg), Class II (pp), and Class III (cc). Class I ADH isozymes oxidize ethanol and other small aliphatic alcohols, and are inhibited by pyrazole. Class II isozymes prefer longer chain aliphatic and aromatic alcohols, are unable to oxidize methanol, and are not inhibited by pyrazole. Class III isozymes prefer even longer chain aliphatic alcohols (five

carbons and longer) and aromatic alcohols, and are not inhibited by pyrazole.

The short-chain alcohol dehydrogenases include a number of related enzymes with a variety of substrate specificities. Included in this group are the mammalian enzymes D-beta-hydroxybutyrate dehydrogenase, (R)-3-hydroxybutyrate dehydrogenase, 15-hydroxyprostaglandin dehydrogenase,

- 5 NADPH-dependent carbonyl reductase, corticosteroid 11-beta-dehydrogenase, and estradiol 17-beta-dehydrogenase, as well as the bacterial enzymes acetoacetyl-CoA reductase, glucose 1-dehydrogenase, 3-beta-hydroxysteroid dehydrogenase, 20-beta-hydroxysteroid dehydrogenase, ribitol dehydrogenase, 3-oxoacyl reductase, 2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase, sorbitol-6-phosphate 2-dehydrogenase, 7-alpha-hydroxysteroid dehydrogenase, cis-1,2-dihydroxy-3,4-
- 10 cyclohexadiene-1-carboxylate dehydrogenase, cis-toluene dihydrodiol dehydrogenase, cis-benzene glycol dehydrogenase, biphenyl-2,3-dihydro-2,3-diol dehydrogenase, N-acylmannosamine 1-dehydrogenase, and 2-deoxy-D-gluconate 3-dehydrogenase (Krozowski, Z. (1994) J. Steroid Biochem. Mol. Biol. 51:125-130; Krozowski, Z. (1992) Mol. Cell Endocrinol. 84:C25-31; and Marks, A.R. et al (1992) J. Biol. Chem. 267:15459-15463).

#### 15 UDP glucuronyltransferase

- Members of the UDP glucuronyltransferase family (UGTs) catalyze the transfer of a glucuronic acid group from the cofactor uridine diphosphate-glucuronic acid (UDP-glucuronic acid) to a substrate. The transfer is generally to a nucleophilic heteroatom (O, N, or S). Substrates include xenobiotics which have been functionalized by Phase I reactions, as well as endogenous compounds
- 20 such as bilirubin, steroid hormones, and thyroid hormones. Products of glucuronidation are excreted in urine if the molecular weight of the substrate is less than about 250 g/mol, whereas larger glucuronidated substrates are excreted in bile.

UGTs are located in the microsomes of liver, kidney, intestine, skin, brain, spleen, and nasal mucosa, where they are on the same side of the endoplasmic reticulum membrane as cytochrome

- 25 P450 enzymes and flavin-containing monooxygenases, and therefore are ideally located to access products of Phase I drug metabolism. UGTs have a C-terminal membrane-spanning domain which anchors them in the endoplasmic reticulum membrane, and a conserved signature domain of about 50 amino acid residues in their C terminal section (Prosite PDOC00359 UDP-glycosyltransferase signature).

- 30 UGTs involved in drug metabolism are encoded by two gene families, UGT1 and UGT2. Members of the UGT1 family result from alternative splicing of a single gene locus, which has a variable substrate binding domain and constant region involved in cofactor binding and membrane insertion. Members of the UGT2 family are encoded by separate gene loci, and are divided into two

families, UGT2A and UGT2B. The 2A subfamily is expressed in olfactory epithelium, and the 2B subfamily is expressed in liver microsomes. Mutations in UGT genes are associated with hyperbilirubinemia (OMIM #143500 Hyperbilirubinemia I); Crigler-Najjar syndrome, characterized by intense hyperbilirubinemia from birth (OMIM #218800 Crigler-Najjar syndrome); and a milder form of

5 hyperbilirubinemia termed Gilbert's disease (OMIM \*191740 UGT1).

#### Sulfotransferase

Sulfate conjugation occurs on many of the same substrates which undergo O-glucuronidation to produce a highly water-soluble sulfuric acid ester. Sulfotransferases (ST) catalyze this reaction by transferring  $\text{SO}_3^-$  from the cofactor 3'-phosphoadenosine-5'-phosphosulfate (PAPS) to the substrate.

10 ST substrates are predominantly phenols and aliphatic alcohols, but also include aromatic amines and aliphatic amines, which are conjugated to produce the corresponding sulfamates. The products of these reactions are excreted mainly in urine.

STs are found in a wide range of tissues, including liver, kidney, intestinal tract, lung, platelets, and brain. The enzymes are generally cytosolic, and multiple forms are often co-expressed. For

15 example, there are more than a dozen forms of ST in rat liver cytosol. These biochemically characterized STs fall into five classes based on their substrate preference: arylsulfotransferase, alcohol sulfotransferase, estrogen sulfotransferase, tyrosine ester sulfotransferase, and bile salt sulfotransferase.

ST enzyme activity varies greatly with sex and age in rats. The combined effects of

20 developmental cues and sex-related hormones are thought to lead to these differences in ST expression profiles, as well as the profiles of other DMEs such as cytochromes P450. Notably, the high expression of STs in cats partially compensates for their low level of UDP glucuronyltransferase activity.

Several forms of ST have been purified from human liver cytosol and cloned. There are two

25 phenol sulfotransferases with different thermal stabilities and substrate preferences. The thermostable enzyme catalyzes the sulfation of phenols such as para-nitrophenol, minoxidil, and acetaminophen; the thermolabile enzyme prefers monoamine substrates such as dopamine, epinephrine, and levadopa. Other cloned STs include an estrogen sulfotransferase and an N-acetylglucosamine-6-O-sulfotransferase. This last enzyme is illustrative of the other major role of STs in cellular biochemistry,

30 the modification of carbohydrate structures that may be important in cellular differentiation and maturation of proteoglycans. Indeed, an inherited defect in a sulfotransferase has been implicated in macular corneal dystrophy, a disorder characterized by a failure to synthesize mature keratan sulfate proteoglycans (Nakazawa, K. et al. (1984) J. Biol. Chem. 259:13751-7; OMIM \*217800 Macular

dystrophy, corneal).

#### Galactosyltransferases

Galactosyltransferases are a subset of glycosyltransferases that transfer galactose (Gal) to the terminal N-acetylglucosamine (GlcNAc) oligosaccharide chains that are part of glycoproteins or glycolipids that are free in solution (Kolbinger, F. et al. (1998) *J. Biol. Chem.* 273:433-440; Amado, M. et al. (1999) *Biochim. Biophys. Acta* 1473:35-53). Galactosyltransferases have been detected on the cell surface and as soluble extracellular proteins, in addition to being present in the Golgi.  $\beta$ 1,3-galactosyltransferases form Type I carbohydrate chains with Gal ( $\beta$ 1-3)GlcNAc linkages. Known human and mouse  $\beta$ 1,3-galactosyltransferases appear to have a short cytosolic domain, a single transmembrane domain, and a catalytic domain with eight conserved regions. (Kolbinger, F. *supra* and Hennot, T. et al. (1998) *J. Biol. Chem.* 273:58-65). In mouse UDP-galactose: $\beta$ -N-acetylglucosamine  $\beta$ 1,3-galactosyltransferase-I region 1 is located at amino acid residues 78-83, region 2 is located at amino acid residues 93-102, region 3 is located at amino acid residues 116-119, region 4 is located at amino acid residues 147-158, region 5 is located at amino acid residues 172-183, region 6 is located at amino acid residues 203-206, region 7 is located at amino acid residues 236-246, and region 8 is located at amino acid residues 264-275. A variant of a sequence found within mouse UDP-galactose: $\beta$ -N-acetylglucosamine  $\beta$ 1,3-galactosyltransferase-I region 8 is also found in bacterial galactosyltransferases, suggesting that this sequence defines a galactosyltransferase sequence motif (Hennot, T. *supra*). Recent work suggests that brainiac protein is a  $\beta$ 1,3-galactosyltransferase. (Yuan, Y. et al. (1997) *Cell* 88:9-11; and Hennot, T. *supra*).

UDP-Gal:GlcNAc-1,4-galactosyltransferase (-1,4-GalT) (Sato, T. et al., (1997) *EMBO J.* 16:1850-1857) catalyzes the formation of Type II carbohydrate chains with Gal ( $\beta$ 1-4)GlcNAc linkages. As is the case with the  $\beta$ 1,3-galactosyltransferase, a soluble form of the enzyme is formed by cleavage of the membrane-bound form. Amino acids conserved among  $\beta$ 1,4-galactosyltransferases include two cysteines linked through a disulfide-bonded and a putative UDP-galactose-binding site in the catalytic domain (Yadav, S. and Brew, K. (1990) *J. Biol. Chem.* 265:14163-14169; Yadav, S.P. and Brew, K. (1991) *J. Biol. Chem.* 266:698-703; and Shaper, N.L. et al. (1997) *J. Biol. Chem.* 272:31389-31399).  $\beta$ 1,4-galactosyltransferases have several specialized roles in addition to synthesizing carbohydrate chains on glycoproteins or glycolipids. In mammals a  $\beta$ 1,4-galactosyltransferase, as part of a heterodimer with  $\alpha$ -lactalbumin, functions in lactating mammary gland lactose production. A  $\beta$ 1,4-galactosyltransferase on the surface of sperm functions as a receptor that specifically recognizes the egg. Cell surface  $\beta$ 1,4-galactosyltransferases also function in cell adhesion, cell/basal lamina interaction, and normal and metastatic cell migration. (Shur,

B. (1993) *Curr. Opin. Cell Biol.* 5:854-863; and Shaper, J. (1995) *Adv. Exp. Med. Biol.* 376:95-104).

#### Glutathione S-transferase

The basic reaction catalyzed by glutathione S-transferases (GST) is the conjugation of an electrophile with reduced glutathione (GSH). GSTs are homodimeric or heterodimeric proteins

5 localized mainly in the cytosol, but some level of activity is present in microsomes as well. The major isozymes share common structural and catalytic properties; in humans they have been classified into four major classes, Alpha, Mu, Pi, and Theta. The two largest classes, Alpha and Mu, are identified by their respective protein isoelectric points; pI ~ 7.5-9.0 (Alpha), and pI ~ 6.6 (Mu). Each GST possesses a common binding site for GSH and a variable hydrophobic binding site. The hydrophobic  
10 binding site in each isozyme is specific for particular electrophilic substrates. Specific amino acid residues within GSTs have been identified as important for these binding sites and for catalytic activity. Residues Q67, T68, D101, E104, and R131 are important for the binding of GSH (Lee, H-C et al. (1995) *J. Biol. Chem.* 270: 99-109). Residues R13, R20, and R69 are important for the catalytic activity of GST (Stenberg G et al. (1991) *Biochem. J.* 274: 549-55).

15 In most cases, GSTs perform the beneficial function of deactivation and detoxification of potentially mutagenic and carcinogenic chemicals. However, in some cases their action is detrimental and results in activation of chemicals with consequent mutagenic and carcinogenic effects. Some forms of rat and human GSTs are reliable preneoplastic markers that aid in the detection of carcinogenesis. Expression of human GSTs in bacterial strains, such as Salmonella typhimurium used  
20 in the well-known Ames test for mutagenicity, has helped to establish the role of these enzymes in mutagenesis. Dihalomethanes, which produce liver tumors in mice, are believed to be activated by GST. This view is supported by the finding that dihalomethanes are more mutagenic in bacterial cells expressing human GST than in untransfected cells (Thier, R. et al. (1993) *Proc. Natl. Acad. Sci. USA* 90: 8567-80). The mutagenicity of ethylene dibromide and ethylene dichloride is increased in bacterial  
25 cells expressing the human Alpha GST, A1-1, while the mutagenicity of aflatoxin B1 is substantially reduced by enhancing the expression of GST (Simula, T.P. et al. (1993) *Carcinogenesis* 14: 1371-6). Thus, control of GST activity may be useful in the control of mutagenesis and carcinogenesis.

GST has been implicated in the acquired resistance of many cancers to drug treatment, the phenomenon known as multi-drug resistance (MDR). MDR occurs when a cancer patient is treated  
30 with a cytotoxic drug such as cyclophosphamide and subsequently becomes resistant to this drug and to a variety of other cytotoxic agents as well. Increased GST levels are associated with some of these drug resistant cancers, and it is believed that this increase occurs in response to the drug agent which is then deactivated by the GST catalyzed GSH conjugation reaction. The increased GST levels

then protect the cancer cells from other cytotoxic agents which bind to GST. Increased levels of A1-1 in tumors has been linked to drug resistance induced by cyclophosphamide treatment (Dirven H.A. et al. (1994) Cancer Res. 54: 6215-20). Thus control of GST activity in cancerous tissues may be useful in treating MDR in cancer patients.

5 Gamma-glutamyl transpeptidase

Gamma-glutamyl transpeptidases are ubiquitously expressed enzymes that initiate extracellular glutathione (GSH) breakdown by cleaving gamma-glutamyl amide bonds. The breakdown of GSH provides cells with a regional cysteine pool for biosynthetic pathways. Gamma-glutamyl transpeptidases also contribute to cellular antioxidant defenses and expression is induced by oxidative  
10 steress. The cell surface-localized glycoproteins are expressed at high levels in cancer cells. Studies have suggested that the high level of gamma-glutamyl transpeptidases activity present on the surface of cancer cells could be exploited to activate precursor drugs, resulting in high local concentrations of anti-cancer therapeutic agents (Hanigan, M.H. (1998) Chem. Biol. Interact. 111-112:333-42; Taniguchi, N. and Ikeda, Y. (1998) Adv. Enzymol. Relat. Areas Mol. Biol. 72:239-78; Chikhi, N. et al.  
15 (1999) Comp. Biochem. Physiol. B. Biochem. Mol. Biol. 122:367-80).

Acyltransferase

N-acyltransferase enzymes catalyze the transfer of an amino acid conjugate to an activated carboxylic group. Endogenous compounds and xenobiotics are activated by acyl-CoA synthetases in the cytosol, microsomes, and mitochondria. The acyl-CoA intermediates are then conjugated with an  
20 amino acid (typically glycine, glutamine, or taurine, but also ornithine, arginine, histidine, serine, aspartic acid, and several dipeptides) by N-acyltransferases in the cytosol or mitochondria to form a metabolite with an amide bond. This reaction is complementary to O-glucuronidation, but amino acid conjugation does not produce the reactive and toxic metabolites which often result from glucuronidation.

One well-characterized enzyme of this class is the bile acid-CoA:amino acid N-acyltransferase  
25 (BAT) responsible for generating the bile acid conjugates which serve as detergents in the gastrointestinal tract (Falany, C. N. et al. (1994) J. Biol. Chem. 269:19375-9; Johnson, M. R. et al. (1991) J. Biol. Chem. 266:10227-33). BAT is also useful as a predictive indicator for prognosis of hepatocellular carcinoma patients after partial hepatectomy (Furutani, M. et al. (1996) Hepatology 24:1441-5).

30 Acetyltransferases

Acetyltransferases have been extensively studied for their role in histone acetylation. Histone acetylation results in the relaxing of the chromatin structure in eukaryotic cells, allowing transcription factors to gain access to promoter elements of the DNA templates in the affected region of the

genome (or the genome in general). In contrast, histone deacetylation results in a reduction in transcription by closing the chromatin structure and limiting access of transcription factors. To this end, a common means of stimulating cell transcription is the use of chemical agents that inhibit the deacetylation of histones (e.g., sodium butyrate), resulting in a global (albeit artifactual) increase in gene expression. The modulation of gene expression by acetylation also results from the acetylation of other proteins, including but not limited to, p53, GATA-1, MyoD, ACTR, TFIIIE, TFIIIF and the high mobility group proteins (HMG). In the case of p53, acetylation results in increased DNA binding, leading to the stimulation of transcription of genes regulated by p53. The prototypic histone acetylase (HAT) is Gcn5 from *Saccharomyces cerevisiae*. Gcn5 is a member of a family of acetylases that includes Tetrahymena p55, human Gcn5, and human p300/CBP. Histone acetylation is reviewed in (Cheung, W.L. et al. (2000) Curr. Opin. Cell Biol. 12:326-333 and Berger, S.L (1999) Curr. Opin. Cell Biol. 11:336-341). Some acetyltransferase enzymes possess the alpha/beta hydrolase fold (Center of Applied Molecular Engineering Inst. of Chemistry and Biochemistry - University of Salzburg, <http://predict.sanger.ac.uk/irbm-course97/Docs/ms/>) common to several other major classes of enzymes, including but not limited to, acetylcholinesterases and carboxylesterases (Structural Classification of Proteins, <http://scop.mrc-lmb.cam.ac.uk/scop/index.html>).

#### N-acetyltransferase

Aromatic amines and hydrazine-containing compounds are subject to N-acetylation by the N-acetyltransferase enzymes of liver and other tissues. Some xenobiotics can be O-acetylated to some extent by the same enzymes. N-acetyltransferases are cytosolic enzymes which utilize the cofactor acetyl-coenzyme A (acetyl-CoA) to transfer the acetyl group in a two step process. In the first step, the acetyl group is transferred from acetyl-CoA to an active site cysteine residue; in the second step, the acetyl group is transferred to the substrate amino group and the enzyme is regenerated.

In contrast to most other DME classes, there are a limited number of known N-acetyltransferases. In humans, there are two highly similar enzymes, NAT1 and NAT2; mice appear to have a third form of the enzyme, NAT3. The human forms of N-acetyltransferase have independent regulation (NAT1 is widely-expressed, whereas NAT2 is in liver and gut only) and overlapping substrate preferences. Both enzymes appear to accept most substrates to some extent, but NAT1 does prefer some substrates (para-aminobenzoic acid, para-aminosalicylic acid, sulfamethoxazole, and sulfanilamide), while NAT2 prefers others (isoniazid, hydralazine, procainamide, dapson, aminogluthethimide, and sulfamethazine).

Clinical observations of patients taking the antituberculosis drug isoniazid in the 1950s led to the description of fast and slow acetylators of the compound. These phenotypes were shown

subsequently to be due to mutations in the NAT2 gene which affected enzyme activity or stability. The slow isoniazid acetylator phenotype is very prevalent in Middle Eastern populations (approx. 70%), and is less prevalent in Caucasian (approx. 50%) and Asian (<25%) populations. More recently, functional polymorphism in NAT1 has been detected, with approximately 8% of the population tested showing a slow acetylator phenotype (Butcher, N. J. et al. (1998) Pharmacogenetics 8:67-72). Since NAT1 can activate some known aromatic amine carcinogens, polymorphism in the widely-expressed NAT1 enzyme may be important in determining cancer risk (OMIM \*108345 N-acetyltransferase 1).

#### Aminotransferases

Aminotransferases comprise a family of pyridoxal 5'-phosphate (PLP) -dependent enzymes that catalyze transformations of amino acids. Aspartate aminotransferase (AspAT) is the most extensively studied PLP-containing enzyme. It catalyzes the reversible transamination of dicarboxylic L-amino acids, aspartate and glutamate, and the corresponding 2-oxo acids, oxalacetate and 2-oxoglutarate. Other members of the family included pyruvate aminotransferase, branched-chain amino acid aminotransferase, tyrosine aminotransferase, aromatic aminotransferase, alanine:glyoxylate aminotransferase (AGT), and kynurenine aminotransferase (Vacca, R.A. et al. (1997) J. Biol. Chem. 272:21932-21937).

Primary hyperoxaluria type-1 is an autosomal recessive disorder resulting in a deficiency in the liver-specific peroxisomal enzyme, alanine:glyoxylate aminotransferase-1. The phenotype of the disorder is a deficiency in glyoxylate metabolism. In the absence of AGT, glyoxylate is oxidized to oxalate rather than being transaminated to glycine. The result is the deposition of insoluble calcium oxalate in the kidneys and urinary tract, ultimately causing renal failure (Lumb, M.J. et al. (1999) J. Biol. Chem. 274:20587-20596).

Kynurenine aminotransferase catalyzes the irreversible transamination of the L-tryptophan metabolite L-kynurenine to form kynurenic acid. The enzyme may also catalyzes the reversible transamination reaction between L-2-aminoadipate and 2-oxoglutarate to produce 2-oxoadipate and L-glutamate. Kynurenic acid is a putative modulator of glutamatergic neurotransmission, thus a deficiency in kynurenine aminotransferase may be associated with pleiotropic effects (Buchli, R. et al. (1995) J. Biol. Chem. 270:29330-29335).

#### Copper-zinc superoxide dismutases

Copper-zinc superoxide dismutases are compact homodimeric metalloenzymes involved in cellular defenses against oxidative damage. The enzymes contain one atom of zinc and one atom of copper per subunit and catalyze the dismutation of superoxide anions into  $O_2$  and  $H_2O_2$ . The rate of

dismutation is diffusion-limited and consequently enhanced by the presence of favorable electrostatic interactions between the substrate and enzyme active site. Examples of this class of enzyme have been identified in the cytoplasm of all the eukaryotic cells as well as in the periplasm of several bacterial species. Copper-zinc superoxide dismutases are robust enzymes that are highly resistant to proteolytic digestion and denaturing by urea and SDS. In addition to the compact structure of the enzymes, the presence of the metal ions and intrasubunit disulfide bonds is believed to be responsible for enzyme stability. The enzymes undergo reversible denaturation at temperatures as high as 70 °C (Battistoni, A. et al. (1998) J. Biol. Chem. 273:5655-5661).

Overexpression of superoxide dismutase has been implicated in enhancing freezing tolerance of transgenic Alfalfa as well as providing resistance to environmental toxins such as the diphenyl ether herbicide, acifluorfen (McKersie, B.D. et al. (1993) Plant Physiol. 103:1155-1163). In addition, yeast cells become more resistant to freeze-thaw damage following exposure to hydrogen peroxide which causes the yeast cells to adapt to further peroxide stress by upregulating expression of superoxide dismutases. In this study, mutations to yeast superoxide dismutase genes had a more detrimental effect on freeze-thaw resistance than mutations which affected the regulation of glutathione metabolism, long suspected of being important in determining an organisms survival through the process of cryopreservation (Jong-In Park, J.-I. et al. (1998) J. Biol. Chem. 273:22921-22928).

Expression of superoxide dismutase is also associated with Mycobacterium tuberculosis, the organism that causes tuberculosis. Superoxide dismutase is one of the ten major proteins secreted by M. tuberculosis and its expression is upregulated approximately 5-fold in response to oxidative stress. M. tuberculosis expresses almost two orders of magnitude more superoxide dismutase than the nonpathogenic mycobacterium M. smegmatis, and secretes a much higher proportion of the expressed enzyme. The result is the secretion of ~350-fold more enzyme by M. tuberculosis than M. smegmatis, providing substantial resistance to oxidative stress (Harth, G. and Horwitz, M.A. (1999) J. Biol. Chem. 274:4281-4292).

The reduced expression of copper-zinc superoxide dismutases, as well as other enzymes with anti-oxidant capabilities, has been implicated in the early stages of cancer. The expression of copper-zinc superoxide dismutases has been shown to be lower in prostatic intraepithelial neoplasia and prostate carcinomas, compared to normal prostate tissue (Bostwick, D.G. (2000) Cancer 89:123-134).

#### Phosphodiesterases

Phosphodiesterases make up a class of enzymes which catalyze the hydrolysis of one of the two ester bonds in a phosphodiester compound. Phosphodiesterases are therefore crucial to a variety of cellular processes. Phosphodiesterases include DNA and RNA endonucleases and exonucleases,

which are essential for cell growth and replication, and topoisomerases, which break and rejoin nucleic acid strands during topological rearrangement of DNA. A Tyr-DNA phosphodiesterase functions in DNA repair by hydrolyzing dead-end covalent intermediates formed between topoisomerase I and DNA (Pouliot, J.J. et al. (1999) *Science* 286:552-555; Yang, S.-W. (1996) *Proc. Natl. Acad. Sci.*

5 USA 93:11534-11539).

Acid sphingomyelinase is a phosphodiesterase which hydrolyzes the membrane phospholipid sphingomyelin to produce ceramide and phosphorylcholine. Phosphorylcholine is used in the synthesis of phosphatidylcholine, which is involved in numerous intracellular signaling pathways, while ceramide is an essential precursor for the generation of gangliosides, membrane lipids found in high  
10 concentration in neural tissue. Defective acid sphingomyelinase leads to a build-up of sphingomyelin molecules in lysosomes, resulting in Niemann-Pick disease (Schuchman, E.H. and S.R. Miranda (1997) *Genet. Test.* 1:13-19).

Glycerophosphoryl diester phosphodiesterase (also known as glycerophosphodiester phosphodiesterase) is a phosphodiesterase which hydrolyzes deacetylated phospholipid  
15 glycerophosphodiesters to produce sn-glycerol-3-phosphate and an alcohol. Glycerophosphocholine, glycerophosphoethanolamine, glycerophosphoglycerol, and glycerophosphoinositol are examples of substrates for glycerophosphoryl diester phosphodiesterases. A glycerophosphoryl diester phosphodiesterase from *E. coli* has broad specificity for glycerophosphodiester substrates (Larson, T.J. et al. (1983) *J. Biol. Chem.* 248:5428-5432).

20 Cyclic nucleotide phosphodiesterases (PDEs) are crucial enzymes in the regulation of the cyclic nucleotides cAMP and cGMP. cAMP and cGMP function as intracellular second messengers to transduce a variety of extracellular signals including hormones, light, and neurotransmitters. PDEs degrade cyclic nucleotides to their corresponding monophosphates, thereby regulating the intracellular concentrations of cyclic nucleotides and their effects on signal transduction. Due to their roles as  
25 regulators of signal transduction, PDEs have been extensively studied as chemotherapeutic targets (Perry, M.J. and G.A. Higgs (1998) *Curr. Opin. Chem. Biol.* 2:472-481; Torphy, J.T. (1998) *Am. J. Resp. Crit. Care Med.* 157:351-370).

Families of mammalian PDEs have been classified based on their substrate specificity and affinity, sensitivity to cofactors, and sensitivity to inhibitory agents (Beavo, J.A. (1995) *Physiol. Rev.*  
30 75:725-748; Conti, M. et al. (1995) *Endocrine Rev.* 16:370-389). Several of these families contain distinct genes, many of which are expressed in different tissues as splice variants. Within PDE families, there are multiple isozymes and multiple splice variants of these isozymes (Conti, M. and S.-L.C. Jin (1999) *Prog. Nucleic Acid Res. Mol. Biol.* 63:1-38). The existence of multiple PDE families,

isozymes, and splice variants is an indication of the variety and complexity of the regulatory pathways involving cyclic nucleotides (Houslay, M.D. and G. Milligan (1997) Trends Biochem. Sci. 22:217-224).

Type 1 PDEs (PDE1s) are  $\text{Ca}^{2+}$ /calmodulin-dependent and appear to be encoded by at least three different genes, each having at least two different splice variants (Kakkar, R. et al. (1999) Cell Mol. Life Sci. 55:1164-1186). PDE1s have been found in the lung, heart, and brain. Some PDE1 isozymes are regulated in vitro by phosphorylation/dephosphorylation. Phosphorylation of these PDE1 isozymes decreases the affinity of the enzyme for calmodulin, decreases PDE activity, and increases steady state levels of cAMP (Kakkar, supra). PDE1s may provide useful therapeutic targets for disorders of the central nervous system, and the cardiovascular and immune systems due to the involvement of PDE1s in both cyclic nucleotide and calcium signaling (Perry, M.J. and G.A. Higgs (1998) Curr. Opin. Chem. Biol. 2:472-481).

PDE2s are cGMP-stimulated PDEs that have been found in the cerebellum, neocortex, heart, kidney, lung, pulmonary artery, and skeletal muscle (Sadhu, K. et al. (1999) J. Histochem. Cytochem. 47:895-906). PDE2s are thought to mediate the effects of cAMP on catecholamine secretion, participate in the regulation of aldosterone (Beavo, supra), and play a role in olfactory signal transduction (Juilfs, D.M. et al. (1997) Proc. Natl. Acad. Sci. USA 94:3388-3395).

PDE3s have high affinity for both cGMP and cAMP, and so these cyclic nucleotides act as competitive substrates for PDE3s. PDE3s play roles in stimulating myocardial contractility, inhibiting platelet aggregation, relaxing vascular and airway smooth muscle, inhibiting proliferation of T-lymphocytes and cultured vascular smooth muscle cells, and regulating catecholamine-induced release of free fatty acids from adipose tissue. The PDE3 family of phosphodiesterases are sensitive to specific inhibitors such as cilostamide, enoximone, and lixazinone. Isozymes of PDE3 can be regulated by cAMP-dependent protein kinase, or by insulin-dependent kinases (Degerman, E. et al. (1997) J. Biol. Chem. 272:6823-6826).

PDE4s are specific for cAMP; are localized to airway smooth muscle, the vascular endothelium, and all inflammatory cells; and can be activated by cAMP-dependent phosphorylation. Since elevation of cAMP levels can lead to suppression of inflammatory cell activation and to relaxation of bronchial smooth muscle, PDE4s have been studied extensively as possible targets for novel anti-inflammatory agents, with special emphasis placed on the discovery of asthma treatments. PDE4 inhibitors are currently undergoing clinical trials as treatments for asthma, chronic obstructive pulmonary disease, and atopic eczema. All four known isozymes of PDE4 are susceptible to the inhibitor rolipram, a compound which has been shown to improve behavioral memory in mice (Barad, M. et al. (1998) Proc. Natl. Acad. Sci. USA 95:15020-15025). PDE4 inhibitors have also been

studied as possible therapeutic agents against acute lung injury, endotoxemia, rheumatoid arthritis, multiple sclerosis, and various neurological and gastrointestinal indications (Doherty, A.M. (1999) *Curr. Opin. Chem. Biol.* 3:466-473).

PDE5 is highly selective for cGMP as a substrate (Turko, I.V. et al. (1998) *Biochemistry* 37:4200-4205), and has two allosteric cGMP-specific binding sites (McAllister-Lucas, L.M. et al. (1995) *J. Biol. Chem.* 270:30671-30679). Binding of cGMP to these allosteric binding sites seems to be important for phosphorylation of PDE5 by cGMP-dependent protein kinase rather than for direct regulation of catalytic activity. High levels of PDE5 are found in vascular smooth muscle, platelets, lung, and kidney. The inhibitor zaprinast is effective against PDE5 and PDE1s. Modification of zaprinast to provide specificity against PDE5 has resulted in sildenafil (VIAGRA; Pfizer, Inc., New York NY), a treatment for male erectile dysfunction (Terrett, N. et al. (1996) *Bioorg. Med. Chem. Lett.* 6:1819-1824). Inhibitors of PDE5 are currently being studied as agents for cardiovascular therapy (Perry, M.J. and G.A. Higgs (1998) *Curr. Opin. Chem. Biol.* 2:472-481).

PDE6s, the photoreceptor cyclic nucleotide phosphodiesterases, are crucial components of the phototransduction cascade. In association with the G-protein transducin, PDE6s hydrolyze cGMP to regulate cGMP-gated cation channels in photoreceptor membranes. In addition to the cGMP-binding active site, PDE6s also have two high-affinity cGMP-binding sites which are thought to play a regulatory role in PDE6 function (Artemyev, N.O. et al. (1998) *Methods* 14:93-104). Defects in PDE6s have been associated with retinal disease. Retinal degeneration in the rd mouse (Yan, W. et al. (1998) *Invest. Ophthalmol. Vis. Sci.* 39:2529-2536), autosomal recessive retinitis pigmentosa in humans (Danciger, M. et al. (1995) *Genomics* 30:1-7), and rod/cone dysplasia 1 in Irish Setter dogs (Suber, M.L. et al. (1993) *Proc. Natl. Acad. Sci. USA* 90:3968-3972) have been attributed to mutations in the PDE6B gene.

The PDE7 family of PDEs consists of only one known member having multiple splice variants (Bloom, T.J. and J.A. Beavo (1996) *Proc. Natl. Acad. Sci. USA* 93:14188-14192). PDE7s are cAMP specific, but little else is known about their physiological function. Although mRNAs encoding PDE7s are found in skeletal muscle, heart, brain, lung, kidney, and pancreas, expression of PDE7 proteins is restricted to specific tissue types (Han, P. et al. (1997) *J. Biol. Chem.* 272:16152-16157; Perry, M.J. and G.A. Higgs (1998) *Curr. Opin. Chem. Biol.* 2:472-481). PDE7s are very closely related to the PDE4 family; however, PDE7s are not inhibited by rolipram, a specific inhibitor of PDE4s (Beavo, *supra*).

PDE8s are cAMP specific, and are closely related to the PDE4 family. PDE8s are expressed in thyroid gland, testis, eye, liver, skeletal muscle, heart, kidney, ovary, and brain. The cAMP-

- hydrolyzing activity of PDE8s is not inhibited by the PDE inhibitors rolipram, vinpocetine, milrinone, IBMX (3-isobutyl-1-methylxanthine), or zaprinast, but PDE8s are inhibited by dipyridamole (Fisher, D.A. et al. (1998) *Biochem. Biophys. Res. Commun.* 246:570-577; Hayashi, M. et al. (1998) *Biochem. Biophys. Res. Commun.* 250:751-756; Soderling, S.H. et al. (1998) *Proc. Natl. Acad. Sci. USA* 95:8991-8996).
- PDE9s are cGMP specific and most closely resemble the PDE8 family of PDEs. PDE9s are expressed in kidney, liver, lung, brain, spleen, and small intestine. PDE9s are not inhibited by sildenafil (VIAGRA; Pfizer, Inc., New York NY), rolipram, vinpocetine, dipyridamole, or IBMX (3-isobutyl-1-methylxanthine), but they are sensitive to the PDE5 inhibitor zaprinast (Fisher, D.A. et al. (1998) *J. Biol. Chem.* 273:15559-15564; Soderling, S.H. et al. (1998) *J. Biol. Chem.* 273:15553-15558).
- PDE10s are dual-substrate PDEs, hydrolyzing both cAMP and cGMP. PDE10s are expressed in brain, thyroid, and testis. (Soderling, S.H. et al. (1999) *Proc. Natl. Acad. Sci. USA* 96:7071-7076; Fujishige, K. et al. (1999) *J. Biol. Chem.* 274:18438-18445; Loughney, K. et al (1999) *Gene* 234:109-117).
- PDEs are composed of a catalytic domain of about 270-300 amino acids, an N-terminal regulatory domain responsible for binding cofactors, and, in some cases, a hydrophilic C-terminal domain of unknown function (Conti, M. and S.-L.C. Jin (1999) *Prog. Nucleic Acid Res. Mol. Biol.* 63:1-38). A conserved, putative zinc-binding motif, HDXXHXGXXN, has been identified in the catalytic domain of all PDEs. N-terminal regulatory domains include non-catalytic cGMP-binding domains in PDE2s, PDE5s, and PDE6s; calmodulin-binding domains in PDE1s; and domains containing phosphorylation sites in PDE3s and PDE4s. In PDE5, the N-terminal cGMP-binding domain spans about 380 amino acid residues and comprises tandem repeats of the conserved sequence motif N(R/K)XnFX<sub>3</sub>DB (McAllister-Lucas, L.M. et al. (1993) *J. Biol. Chem.* 268:22863-22873). The NKXnD motif has been shown by mutagenesis to be important for cGMP binding (Turko, I.V. et al. (1996) *J. Biol. Chem.* 271:22240-22244). PDE families display approximately 30% amino acid identity within the catalytic domain; however, isozymes within the same family typically display about 85-95% identity in this region (e.g. PDE4A vs PDE4B). Furthermore, within a family there is extensive similarity (>60%) outside the catalytic domain; while across families, there is little or no sequence similarity outside this domain.
- Many of the constituent functions of immune and inflammatory responses are inhibited by agents that increase intracellular levels of cAMP (Verghese, M.W. et al. (1995) *Mol. Pharmacol.* 47:1164-1171). A variety of diseases have been attributed to increased PDE activity and associated with decreased levels of cyclic nucleotides. For example, a form of diabetes insipidus in mice has

been associated with increased PDE4 activity, an increase in low- $K_m$  cAMP PDE activity has been reported in leukocytes of atopic patients, and PDE3 has been associated with cardiac disease.

Many inhibitors of PDEs have been identified and have undergone clinical evaluation (Perry, M.J. and G.A. Higgs (1998) *Curr. Opin. Chem. Biol.* 2:472-481; Torphy, T.J. (1998) *Am. J. Respir.*

- 5 *Crit. Care Med.* 157:351-370). PDE3 inhibitors are being developed as antithrombotic agents, antihypertensive agents, and as cardiotonic agents useful in the treatment of congestive heart failure. Rolipram, a PDE4 inhibitor, has been used in the treatment of depression, and other inhibitors of PDE4 are undergoing evaluation as anti-inflammatory agents. Rolipram has also been shown to inhibit lipopolysaccharide (LPS) induced TNF- $\alpha$  which has been shown to enhance HIV-1 replication in vitro.  
10 Therefore, rolipram may inhibit HIV-1 replication (Angel, J.B. et al. (1995) *AIDS* 9:1137-1144). Additionally, rolipram, based on its ability to suppress the production of cytokines such as TNF- $\alpha$  and  $\gamma$  and interferon  $\gamma$ , has been shown to be effective in the treatment of encephalomyelitis. Rolipram may also be effective in treating tardive dyskinesia and was effective in treating multiple sclerosis in an experimental animal model (Sommer, N. et al. (1995) *Nat. Med.* 1:244-248; Sasaki, H. et al. (1995)  
15 *Eur. J. Pharmacol.* 282:71-76).

- Theophylline is a nonspecific PDE inhibitor used in the treatment of bronchial asthma and other respiratory diseases. Theophylline is believed to act on airway smooth muscle function and in an anti-inflammatory or immunomodulatory capacity in the treatment of respiratory diseases (Banner, K.H. and C.P. Page (1995) *Eur. Respir. J.* 8:996-1000). Pentoxifylline is another nonspecific PDE  
20 inhibitor used in the treatment of intermittent claudication and diabetes-induced peripheral vascular disease. Pentoxifylline is also known to block TNF- $\alpha$  production and may inhibit HIV-1 replication (Angel et al., supra).

- PDEs have been reported to affect cellular proliferation of a variety of cell types (Conti et al. (1995) *Endocrine Rev.* 16:370-389) and have been implicated in various cancers. Growth of prostate  
25 carcinoma cell lines DU145 and LNCaP was inhibited by delivery of cAMP derivatives and PDE inhibitors (Bang, Y.J. et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:5330-5334). These cells also showed a permanent conversion in phenotype from epithelial to neuronal morphology. It has also been suggested that PDE inhibitors have the potential to regulate mesangial cell proliferation (Matousovich, K. et al. (1995) *J. Clin. Invest.* 96:401-410) and lymphocyte proliferation (Joulain, C. et al. (1995) *J.*  
30 *Lipid Mediat. Cell Signal.* 11:63-79). A cancer treatment has been described that involves intracellular delivery of PDEs to particular cellular compartments of tumors, resulting in cell death (Deonarain, M.P. and A.A. Epenetos (1994) *Br. J. Cancer* 70:786-794).

#### Phosphotriesterases

Phosphotriesterases (PTE, paraoxonases) are enzymes that hydrolyze toxic organophosphorus compounds and have been isolated from a variety of tissues. The enzymes appear to be lacking in birds and insects and abundant in mammals, explain the reduced tolerance of birds and insects to organophosphorus compound (Vilanova, E. and Sogorb, M.A. (1999) Crit. Rev. Toxicol. 29:21-57).

5 Phosphotriesterases play a central role in the detoxification of insecticides by mammals.

Phosphotriesterase activity varies among individuals and is lower in infants than adults. Knockout mice are markedly more sensitive to the organophosphate-based toxins diazoxon and chlorpyrifos oxon (Furlong, C.E., et al. (2000) Neurotoxicology 21:91-100). PTEs have attracted interest as enzymes capable of the detoxification of organophosphate-containing chemical waste and warfare reagents

10 (e.g., parathion), in addition to pesticides and insecticides. Some studies have also implicated phosphotriesterase in atherosclerosis and diseases involving lipoprotein metabolism.

#### Thioesterases

Two soluble thioesterases involved in fatty acid biosynthesis have been isolated from mammalian tissues, one which is active only toward long-chain fatty-acyl thioesters and one which is

15 active toward thioesters with a wide range of fatty-acyl chain-lengths. These thioesterases catalyze the chain-terminating step in the *de novo* biosynthesis of fatty acids. Chain termination involves the hydrolysis of the thioester bond which links the fatty acyl chain to the 4'-phosphopantetheine prosthetic group of the acyl-carrier protein (ACP) subunit of the fatty acid synthase (Smith, S. (1981a) Methods Enzymol. 71:181-188; Smith, S. (1981b) Methods Enzymol. 71:188-200).

20 E. coli contains two soluble thioesterases, thioesterase I which is active only toward long-chain acyl thioesters, and thioesterase II (TEII) which has a broad chain-length specificity (Naggert, J. et al. (1991) J. Biol. Chem. 266:11044-11050). E. coli TEII does not exhibit sequence similarity with either of the two types of mammalian thioesterases which function as chain-terminating enzymes in *de novo* fatty acid biosynthesis. Unlike the mammalian thioesterases, E. coli TEII lacks the characteristic

25 serine active site gly-X-ser-X-gly sequence motif and is not inactivated by the serine modifying agent diisopropyl fluorophosphate. However, modification of histidine 58 by iodoacetamide and diethylpyrocarbonate abolished TEII activity. Overexpression of TEII did not alter fatty acid content in E. coli, which suggests that it does not function as a chain-terminating enzyme in fatty acid biosynthesis (Naggert et al., *supra*). For that reason, Naggert et al. (*supra*) proposed that the

30 physiological substrates for E. coli TEII may be coenzyme A (CoA)-fatty acid esters instead of ACP-phosphopantetheine-fatty acid esters.

#### Carboxylesterases

Mammalian carboxylesterases constitute a multigene family expressed in a variety of tissues

and cell types. Isozymes have significant sequence homology and are classified primarily on the basis of amino acid sequence. Acetylcholinesterase, butyrylcholinesterase, and carboxylesterase are grouped into the serine super family of esterases (B-esterases). Other carboxylesterases included thyroglobulin, thrombin, Factor IX, gliotactin, and plasminogen. Carboxylesterases catalyze the hydrolysis of ester- and amide- groups from molecules and are involved in detoxification of drugs, environmental toxins, and carcinogens. Substrates for carboxylesterases include short- and long-chain acyl-glycerols, acylcarnitine, carbonates, dipivefrin hydrochloride, cocaine, salicylates, capsaicin, palmitoyl-coenzyme A, imidapril, haloperidol, pyrrolizidine alkaloids, steroids, p-nitrophenyl acetate, malathion, butanilcaine, and isocarboxazide. The enzymes often demonstrate low substrate specificity. Carboxylesterases are also important for the conversion of prodrugs to their respective free acids, which may be the active form of the drug (e.g., lovastatin, used to lower blood cholesterol) (reviewed in Satoh, T. and Hosokawa, M. (1998) *Annu. Rev. Pharmacol. Toxicol.* 38:257-288).

Neuroligins are a class of molecules that (i) have N-terminal signal sequences, (ii) resemble cell-surface receptors, (iii) contain carboxylesterase domains, (iv) are highly expressed in the brain, and (v) bind to neurexins in a calcium-dependent manner. Despite the homology to carboxylesterases, neuroligins lack the active site serine residue, implying a role in substrate binding rather than catalysis (Ichitchenko, K. et al. (1996) *J. Biol. Chem.* 271:2676-2682).

#### Squalene epoxidase

Squalene epoxidase (squalene monooxygenase, SE) is a microsomal membrane-bound, FAD-dependent oxidoreductase that catalyzes the first oxygenation step in the sterol biosynthetic pathway of eukaryotic cells. Cholesterol is an essential structural component of cytoplasmic membranes acquired via the LDL receptor-mediated pathway or the biosynthetic pathway. In the latter case, all 27 carbon atoms in the cholesterol molecule are derived from acetyl-CoA (Stryer, L., *supra*). SE converts squalene to 2,3(*S*)-oxidosqualene, which is then converted to lanosterol and then cholesterol. The steps involved in cholesterol biosynthesis are summarized below (Stryer, L (1988) *Biochemistry*. W.H Freeman and Co., Inc. New York. pp. 554-560 and Sakakibara, J. et al. (1995) 270:17-20): acetate (from Acetyl-CoA) → 3-hydroxy-3-methyl-glutaryl CoA → mevalonate → 5-phosphomevalonate → 5-pyrophosphomevalonate → isopentenyl pyrophosphate → dimethylallyl pyrophosphate → geranyl pyrophosphate → farnesyl pyrophosphate → squalene → squalene epoxide → lanosterol → cholesterol

While cholesterol is essential for the viability of eukaryotic cells, inordinately high serum cholesterol levels results in the formation of atherosclerotic plaques in the arteries of higher organisms. This deposition of highly insoluble lipid material onto the walls of essential blood vessels (e.g., coronary arteries) results in decreased blood flow and potential necrosis of the tissues deprived of adequate

blood flow. HMG-CoA reductase is responsible for the conversion of 3-hydroxyl-3-methyl-glutaryl CoA (HMG-CoA) to mevalonate, which represents the first committed step in cholesterol biosynthesis. HMG-CoA is the target of a number of pharmaceutical compounds designed to lower plasma cholesterol levels. However, inhibition of MHG-CoA also results in the reduced synthesis of non-sterol intermediates (e.g., mevalonate) required for other biochemical pathways. SE catalyzes a rate-limiting reaction that occurs later in the sterol synthesis pathway and cholesterol in the only end product of the pathway following the step catalyzed by SE. As a result, SE is the ideal target for the design of anti-hyperlipidemic drugs that do not cause a reduction in other necessary intermediates (Nakamura, Y. et al. (1996) 271:8053-8056).

#### 10 Epoxide hydrolases

Epoxide hydrolases catalyze the addition of water to epoxide-containing compounds, thereby hydrolyzing epoxides to their corresponding 1,2-diols. They are related to bacterial haloalkane dehalogenases and show sequence similarity to other members of the  $\alpha/\beta$  hydrolase fold family of enzymes (e.g., bromoperoxidase A2 from *Streptomyces aureofaciens*, hydroxymuconic semialdehyde hydrolases from *Pseudomonas putida*, and haloalkane dehalogenase from *Xanthobacter autotrophicus*). Epoxide hydrolases are ubiquitous in nature and have been found in mammals, invertebrates, plants, fungi, and bacteria. This family of enzymes is important for the detoxification of xenobiotic epoxide compounds which are often highly electrophilic and destructive when introduced into an organism. Examples of epoxide hydrolase reactions include the hydrolysis of cis-9,10-epoxyoctadec-9(Z)-enoic acid (leukotoxin) to form its corresponding diol, threo-9,10-dihydroxyoctadec-12(Z)-enoic acid (leukotoxin diol), and the hydrolysis of cis-12,13-epoxyoctadec-9(Z)-enoic acid (isoleukotoxin) to form its corresponding diol threo-12,13-dihydroxyoctadec-9(Z)-enoic acid (isoleukotoxin diol). Leukotoxins alter membrane permeability and ion transport and cause inflammatory responses. In addition, epoxide carcinogens are known to be produced by cytochrome P450 as intermediates in the detoxification of drugs and environmental toxins.

The enzymes possess a catalytic triad composed of Asp (the nucleophile), Asp (the histidine-supporting acid), and His (the water-activating histidine). The reaction mechanism of epoxide hydrolase proceeds via a covalently bound ester intermediate initiated by the nucleophilic attack of one of the Asp residues on the primary carbon atom of the epoxide ring of the target molecule, leading to a covalently bound ester intermediate (Michael Arand, M. et al. (1996) J. Biol. Chem. 271:4223-4229; Rink, R. et al. (1997) J. Biol. Chem. 272:14650-14657; Argiriadi, M.A. et al. (2000) J. Biol. Chem. 275:15265-15270).

Catechol-O-methyltransferase:

Catechol-O-methyltransferase (COMT) catalyzes the transfer of the methyl group of S-adenosyl-L-methionine (AdoMet; SAM) donor to one of the hydroxyl groups of the catechol substrate (e.g., L-dopa, dopamine, or DBA). Methylation of the 3'-hydroxyl group is favored over methylation of the 4'-hydroxyl group and the membrane bound isoform of COMT is more regiospecific than the soluble form. Translation of the soluble form of the enzyme results from utilization of an internal start codon in a full-length mRNA (1.5 kb) or from the translation of a shorter mRNA (1.3 kb), transcribed from an internal promoter. The proposed  $S_N2$ -like methylation reaction requires  $Mg^{++}$  and is inhibited by  $Ca^{++}$ . The binding of the donor and substrate to COMT occurs sequentially. AdoMet first binds COMT in a  $Mg^{++}$ -independent manner, followed by the binding of  $Mg^{++}$  and the binding of the catechol substrate.

The amount of COMT in tissues is relatively high compared to the amount of activity normally required, thus inhibition is problematic. Nonetheless, inhibitors have been developed for in vitro use (e.g., gallates, tropolone, U-0521, and 3',4'-dihydroxy-2-methyl-propionphenetropolone) and for clinical use (e.g., nitrocatechol-based compounds and tolcapone). Administration of these inhibitors results in the increased half-life of L-dopa and the consequent formation of dopamine. Inhibition of COMT is also likely to increase the half-life of various other catechol-structure compounds, including but not limited to epinephrine/norepinephrine, isoprenaline, rimeterol, dobutamine, fenoldopam, apomorphine, and  $\alpha$ -methyldopa. A deficiency in norepinephrine has been linked to clinical depression, hence the use of COMT inhibitors could be useful in the treatment of depression. COMT inhibitors are generally well tolerated with minimal side effects and are ultimately metabolized in the liver with only minor accumulation of metabolites in the body (Männistö, P.T. and Kaakkola, S. (1999) Pharmacological Reviews 51:593-628).

The discovery of new drug metabolizing enzymes, and the polynucleotides encoding them, satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of autoimmune/inflammatory, cell proliferative, developmental, endocrine, eye, metabolic, and gastrointestinal disorders, including liver disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of drug metabolizing enzymes.

**SUMMARY OF THE INVENTION**

The invention features purified polypeptides, drug metabolizing enzymes, referred to

collectively as "DME" and individually as "DME-1," "DME-2," "DME-3," "DME-4," "DME-5," "DME-6," "DME-7," "DME-8," "DME-9," "DME-10," "DME-11," "DME-12," "DME-13," "DME-14," "DME-15," "DME-16," "DME-17," "DME-18," and "DME-19." In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-19.

The invention further provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-19. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:20-38.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting

of SEQ ID NO:1-19 , b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-19 , c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19 , and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19 . The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-19 , b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-19 , c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19 , and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19 .

The invention further provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:20-38, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:20-38, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:20-38, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:20-38, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said

probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said  
5 target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:20-38, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:20-38, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the  
10 polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide  
15 selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-19 , b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-19 , c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19 , and d) an immunogenic  
20 fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19 , and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-19 . The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional DMB, comprising administering to a patient in need of such treatment the  
25 composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-19 , b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected  
30 from the group consisting of SEQ ID NO:1-19 , c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19 , and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19 . The method comprises a) exposing a sample comprising the

polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional DME, comprising  
5 administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-19 , b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid  
10 sequence selected from the group consisting of SEQ ID NO:1-19 , c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19 , and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19 . The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the  
15 invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional DME, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to  
20 a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-19 , b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-19 , c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19 , and d) an  
25 immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19 . The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the  
30 activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-19 , b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-19 , c) a biologically active fragment of a polypeptide

having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19 , and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19 . The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:20-38, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:20-38, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:20-38, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:20-38, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:20-38, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of

toxicity of the test compound.

### BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for polypeptides of the invention. The probability score for the match between each polypeptide and its GenBank homolog is also shown.

Table 3 shows structural features of polypeptide sequences of the invention, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

### DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs.

Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

## DEFINITIONS

"DME" refers to the amino acid sequences of substantially purified DME obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of DME. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of DME either by directly interacting with DME or by acting on components of the biological pathway in which DME participates.

An "allelic variant" is an alternative form of the gene encoding DME. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding DME include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as DME or a polypeptide with at least one functional characteristic of DME. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding DME, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding DME. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent DME. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of DME is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and

arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

5       The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

10       "Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

      The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of DME. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small  
15       molecules, or any other compound or composition which modulates the activity of DME either by directly interacting with DME or by acting on components of the biological pathway in which DME participates.

      The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')<sub>2</sub>, and Fv fragments, which are capable of binding an epitopic determinant.  
20       Antibodies that bind DME polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and  
25       keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

      The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on  
30       the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

      The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA;

peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic DME, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding DME or fragments of DME may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least

interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

5	Original Residue	Conservative Substitution
	Ala	Gly, Ser
	Arg	His, Lys
	Asn	Asp, Gln, His
	Asp	Asn, Glu
10	Cys	Ala, Ser
	Gln	Asn, Glu, His
	Glu	Asp, Gln, His
	Gly	Ala
	His	Asn, Arg, Gln, Glu
15	Ile	Leu, Val
	Leu	Ile, Val
	Lys	Arg, Gln, Glu
	Met	Leu, Ile
	Phe	His, Met, Leu, Trp, Tyr
20	Ser	Cys, Thr
	Thr	Ser, Val
	Trp	Phe, Tyr
	Tyr	His, Phe, Trp
25	Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

30 A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide.

Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains  
 35 at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

40 "Differential expression" refers to increased or upregulated; or decreased, downregulated, or

absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

“Exon shuffling” refers to the recombination of different coding regions (exons). Since an  
 5 exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the evolution of new protein functions.

A “fragment” is a unique portion of DME or the polynucleotide encoding DME which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up  
 10 to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For  
 15 example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

20 A fragment of SEQ ID NO:20-38 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:20-38, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:20-38 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:20-38 from related polynucleotide sequences. The precise length of a fragment of SEQ ID  
 25 NO:20-38 and the region of SEQ ID NO:20-38 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-19 is encoded by a fragment of SEQ ID NO:20-38. A fragment of SEQ ID NO:1-19 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-19. For example, a fragment of SEQ ID NO:1-19 is useful as an  
 30 immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-19. The precise length of a fragment of SEQ ID NO:1-19 and the region of SEQ ID NO:1-19 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at <http://www.ncbi.nlm.nih.gov/BLAST/>. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

*Matrix: BLOSUM62*

*Reward for match: 1*

*Penalty for mismatch: -2*

*Open Gap: 5 and Extension Gap: 2 penalties*

*Gap x drop-off: 50*

5 *Expect: 10*

*Word Size: 11*

*Filter: on*

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

15 Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

25 Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version

2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

*Matrix: BLOSUM62*

*Open Gap: 11 and Extension Gap: 1 penalties*

5 *Gap x drop-off: 50*

*Expect: 10*

*Word Size: 3*

*Filter: on*

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

15 "Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic strength and pH. The  $T_m$  is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating  $T_m$  and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g.,  $C_0t$  or  $R_0t$  analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of DME which is capable of eliciting an immune response when introduced into a living organism, for example, a

mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of DME which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

5 The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of DME. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of DME.

10 The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a  
15 functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which  
20 comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an DME may involve lipidation, glycosylation,  
25 phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of DME.

"Probe" refers to nucleic acid sequences encoding DME, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are  
30 isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target

DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also  
5 be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for  
10 example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that  
15 purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000  
20 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer  
25 selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The  
30 PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved

oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be used to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing DME, nucleic acids encoding DME, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a

protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro

fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation.

- 5 Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), *supra*.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-  
10 1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have  
15 significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each  
20 other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

25 A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least  
30 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

## THE INVENTION

The invention is based on the discovery of new human drug metabolizing enzymes (DME), the polynucleotides encoding DME, and the use of these compositions for the diagnosis, treatment, or prevention of autoimmune/inflammatory, cell proliferative, developmental, endocrine, eye, metabolic, and gastrointestinal disorders, including liver disorders.

5 Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is  
10 denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte  
15 polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (Genbank ID NO:) of the nearest GenBank homolog. Column 4 shows the probability score for the match between each polypeptide and its GenBank homolog. Column 5 shows the annotation of the GenBank homolog along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

20 Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS  
25 program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these  
30 properties establish that the claimed polypeptides are drug metabolizing enzymes. For example, SEQ ID NO:3 is 40% identical to a mouse cytochrome P450 monooxygenase (GenBank ID g2653663) as determined by the Basic Local Alignment Search Tool (BLAST, see Table 2). The BLAST probability score is  $5.3e-91$ , which indicates the probability of obtaining the observed polypeptide

sequence alignment by chance. SEQ ID NO:3 also contains cytochrome P450 signature sequences as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains (see Table 3). Data from BLIMPS and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:3 is a member of the cytochrome P450 family.

In an alternative example, SEQ ID NO:1 is 58% identical to a lysyl oxidase from the yellow perch (*Perca flavescens*; GenBank ID g4929199) as determined by BLAST analysis. The BLAST probability score is  $1.9\text{e-}248$ . SEQ ID NO:1 also contains cytochrome P450 signature sequences as determined by searching for statistically significant matches in the HMM-based PFAM database of conserved protein family domains and by BLIMPS analyses.

In an alternative example, SEQ ID NO:2 is 61% identical to human flavin-containing monooxygenase 5 (GenBank ID g559046) as determined by BLAST analysis, with a probability score of  $4.5\text{e-}181$ . SEQ ID NO:2 also contains flavin-containing monooxygenase signature sequences as determined by searching for statistically significant matches in the HMM-based PFAM database of conserved protein family domains and by BLIMPS and PROFILESCAN analyses.

In an alternative example, SEQ ID NO:4 is 39% identical to a *Pseudomonas* 2,3-butanediol dehydrogenase (GenBank ID g529564) as determined by BLAST analysis, with a probability score of  $2.0\text{e-}61$ . SEQ ID NO:4 also contains dehydrogenase signature sequences as determined by searching for statistically significant matches in the HMM-based PFAM database of conserved protein family domains. Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:4 is a dehydrogenase.

SEQ ID NO:5 is 54% identical to a *Bacillus* quinone oxidase (GenBank ID g2633069) as determined by BLAST analysis, with a probability score of  $7.1\text{e-}96$ . Data obtained by searching the HMM-based PFAM database of conserved protein family domains and by BLIMPS analyses provide further corroborative evidence that SEQ ID NO:5 is a quinone oxidase.

In an alternative example, SEQ ID NO:6 is 92% identical to mouse heparan sulfate 6-sulfotransferase 2 (GenBank ID g6683558) as determined by BLAST analysis, with a probability score of  $2.3\text{e-}255$ .

In an alternative example, SEQ ID NO:7 is 90% identical to a human glutathione S-transferase subunit (GenBank ID g242749) as determined by BLAST analysis, with a probability score of  $1.3\text{e-}101$ . SEQ ID NO:7 also contains glutathione S-transferase signature sequences as determined by searching for statistically significant matches in the HMM-based PFAM database of conserved protein family domains and by BLIMPS analyses.

In an alternative example, SEQ ID NO:8 is 40% identical to a human steroid dehydrogenase (GenBank ID g5531815) as determined by BLAST analysis, with a probability score of  $1.9e-56$ . SEQ ID NO:8 also contains dehydrogenase signature sequences as determined by searching for statistically significant matches in the HMM-based PFAM database of conserved protein family domains.

5 SEQ ID NO:9 is 47% identical to a rabbit liver carboxylesterase (GenBank ID g3219695) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is  $6.3e-72$ , which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:9 also contains carboxylesterase domains as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM  
10 database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:9 is a carboxylesterase.

SEQ ID NO:10 is 45% identical to human carboxylesterase (GenBank ID g180950) as determined by the Basic Local Alignment Search Tool (BLAST, see Table 2). The BLAST  
15 probability score is  $8.7e-130$ , which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:10 also contains carboxylesterase domains as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains (see Table 3). Data from BLIMPS,  
MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:10  
20 is a carboxylesterase.

In an alternative example, SEQ ID NO:11 is 89% identical to murine heparan sulfate 6-sulfotransferase 2 (GenBank ID g6683558) as determined by the Basic Local Alignment Search Tool (BLAST, see Table 2). The BLAST probability score is  $1.8e-236$ , which indicates the probability of obtaining the observed polypeptide sequence alignment by chance, and provides evidence that SEQ ID  
25 NO:11 is a DMB, and specifically that SEQ ID NO:11 is a sulfotransferase.

In an alternative example, SEQ ID NO:12 is 25% identical to a Bacillus subtilis epoxide hydrolase (GenBank ID g2633182) as determined by the Basic Local Alignment Search Tool (BLAST, see Table 2). The BLAST probability score is  $1.3e-11$ , which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:12 also contains  
30 hydrolase domains as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains (see Table 3). Data from BLIMPS analyses provide further corroborative evidence that SEQ ID NO:12 is a hydrolase.

In an alternative example, SEQ ID NO:13 is 83% identical to a rat beta-alanine-pyruvate aminotransferase (GenBank ID g1944136) as determined by the BLAST analysis (see Table 2). The BLAST probability score is  $1.1\text{e-}234$ . SEQ ID NO:13 also contains aminotransferase domains as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains (see Table 3). Data from BLIMPS and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:13 is an aminotransferase.

In an alternative example, SEQ ID NO:14 is 50% identical to a guinea pig hydroxysteroid sulfotransferase (GenBank ID g1151081) as determined by the BLAST analysis (see Table 2). The BLAST probability score is  $5.4\text{e-}34$ , and provides evidence that SEQ ID NO:14 is a sulfotransferase.

In an alternative example, SEQ ID NO:15 is 52% identical to a guinea pig copper/zinc superoxide dismutase (GenBank ID g1066120) as determined by the BLAST analysis (see Table 2). The BLAST probability score is  $2.1\text{e-}25$ . SEQ ID NO:15 also contains copper/zinc superoxide dismutase domains as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains (see Table 3). Data from BLIMPS analyses provide further corroborative evidence that SEQ ID NO:15 is a copper/zinc superoxide dismutase.

SEQ ID NO:16 is 37% identical to human 3'-phosphoadenylylsulfate-galactosylceramide 3'-sulfotransferase (cerebroside sulfotransferase, GenBank ID g1871141) as determined by the Basic Local Alignment Search Tool (BLAST, see Table 2). The BLAST probability score is  $2.8\text{e-}60$ , which indicates the probability of obtaining the observed polypeptide sequence alignment by chance.

In an alternative example, SEQ ID NO:17 is 38% identical to a putative *C. elegans* monoamine oxidase (GenBank ID g6782275) as determined by BLAST analysis with a probability score of  $3.0\text{e-}99$ . SEQ ID NO:17 also contains a monoamine oxidase domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains (see Table 3). Data from BLIMPS analysis provide further corroborative evidence that SEQ ID NO:17 is a monoamine oxidase.

In an alternative example, SEQ ID NO:18 is 36% identical to human catechol-O-methyltransferase (GenBank ID g179955) as determined by BLAST analysis with a probability score of  $9.5\text{e-}41$ . SEQ ID NO:18 is also 36% identical to murine catechol-O-methyltransferase (GenBank ID g3493253) as determined by BLAST analysis with a probability score of  $1.3\text{e-}41$ .

In an alternative example, SEQ ID NO:19 is 44% identical to *Fundulus heteroclitus* cytochrome P450 2N1 (GenBank ID g5852342) as determined by the Basic Local Alignment Search

Tool (BLAST, see Table 2). The BLAST probability score is  $4.6e-99$ , which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:19 also contains cytochrome P450 domains as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains (see Table 3). Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:19 is a cytochrome P450.

The algorithms and parameters for the analysis of SEQ ID NO:1-19 are described in Table

7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Columns 1 and 2 list the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and the corresponding Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) for each polynucleotide of the invention. Column 3 shows the length of each polynucleotide sequence in basepairs. Column 4 lists fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:20-38 or that distinguish between SEQ ID NO:20-38 and related polynucleotide sequences. Column 5 shows identification numbers corresponding to cDNA sequences, coding sequences (exons) predicted from genomic DNA, and/or sequence assemblages comprised of both cDNA and genomic DNA. These sequences were used to assemble the full length polynucleotide sequences of the invention. Columns 6 and 7 of Table 4 show the nucleotide start (5') and stop (3') positions of the cDNA and/or genomic sequences in column 5 relative to their respective full length sequences.

The identification numbers in Column 5 of Table 4 may refer specifically, for example, to Incyte cDNAs along with their corresponding cDNA libraries. For example, 7690384J1 is the identification number of an Incyte cDNA sequence, and PROSTME06 is the cDNA library from which it is derived. Incyte cDNAs for which cDNA libraries are not indicated were derived from pooled cDNA libraries (e.g., 55017748J1). Alternatively, the identification numbers in column 5 may refer to GenBank cDNAs or ESTs (e.g., g1203094) which contributed to the assembly of the full length polynucleotide sequences. In addition, the identification numbers in column 5 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the identification numbers in column 5 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records

(i.e., those sequences including the designation "NP"). Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, FL\_XXXXXX\_N<sub>1</sub>\_N<sub>2</sub>\_YYYY\_N<sub>3</sub>\_N<sub>4</sub> represents a "stitched" sequence in which XXXXXX is the identification number of the cluster of sequences to which the algorithm was applied, and YYYY is the number of the prediction generated by the algorithm, and N<sub>1,2,3,4</sub>, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the identification numbers in column 5 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example, FLXXXXXX\_gAAAAA\_gBBBBB\_1\_N is the identification number of a "stretched" sequence, with XXXXXX being the Incyte project identification number, gAAAAA being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, gBBBBB being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and N referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the "exon-stretching" algorithm, a RefSeq identifier (denoted by "NM," "NP," or "NT") may be used in place of the GenBank identifier (i.e., gBBBBB).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

Prefix	Type of analysis and/or examples of programs
GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK)
GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).
INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in column 5 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and  
5 vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

The invention also encompasses DME variants. A preferred DME variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the DME amino acid sequence, and which contains at least one functional or structural characteristic of DME.

10 The invention also encompasses polynucleotides which encode DME. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:20-38, which encodes DME. The polynucleotide sequences of SEQ ID NO:20-38, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar  
15 backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding DME. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding DME. A particular aspect of the invention encompasses a variant of a  
20 polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:20-38 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:20-38. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of DME.

25 It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding DME, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in  
30 accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring DME, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode DME and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring DME under appropriately selected

conditions of stringency, it may be advantageous to produce nucleotide sequences encoding DME or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding DME and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode DME and DME derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding DME or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:20-38 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding DME may be extended utilizing a partial nucleotide

sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.)

- 5 Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060.) Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 10 68°C to 72°C.

- When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

- Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments

which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode DME may be cloned in recombinant DNA molecules that direct expression of DME, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express DME.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter DME-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of DME, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding DME may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, DME itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g.,

Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) *Science* 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of DMB, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) *Methods Enzymol.* 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, *supra*, pp. 28-53.)

In order to express a biologically active DMB, the nucleotide sequences encoding DMB or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding DMB. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding DMB. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding DMB and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding DMB and appropriate transcriptional and translational control elements. These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences

encoding DME. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, *supra*; Ausubel, *supra*; Van Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509; Engelhard, E.K. et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:3224-3227; Sandig, V. et al. (1996) *Hum. Gene Ther.* 7:1937-1945; Takamatsu, N. (1987) *EMBO J.* 6:307-311; *The McGraw Hill Yearbook of Science and Technology* (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:3655-3659; and Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) *Cancer Gen. Ther.* 5(6):350-356; Yu, M. et al. (1993) *Proc. Natl. Acad. Sci. USA* 90(13):6340-6344; Buller, R.M. et al. (1985) *Nature* 317(6040):813-815; McGregor, D.P. et al. (1994) *Mol. Immunol.* 31(3):219-226; and Verma, I.M. and N. Somia (1997) *Nature* 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding DME. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding DME can be achieved using a multifunctional *E. coli* vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding DME into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for *in vitro* transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509.) When large quantities of DME are needed, e.g. for the production of antibodies, vectors which direct high level expression of DME may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of DME. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast *Saccharomyces cerevisiae* or *Pichia pastoris*. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration

of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, G.A. et al. (1987) *Methods Enzymol.* 153:516-544; and Scorer, C.A. et al. (1994) *Bio/Technology* 12:181-184.)

Plant systems may also be used for expression of DME. Transcription of sequences encoding  
5 DME may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) *EMBO J.* 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) *EMBO J.* 3:1671-1680; Broglie, R. et al. (1984) *Science* 224:838-843; and Winter, J. et al. (1991) *Results Probl. Cell Differ.* 17:85-105.) These  
10 constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding DME may be ligated into  
15 an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses DME in host cells. (See, e.g., Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-  
20 based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-  
25 355.)

For long term production of recombinant proteins in mammalian systems, stable expression of DME in cell lines is preferred. For example, sequences encoding DME can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the  
30 introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue

culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk* and *ap<sup>r</sup>* cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech),  $\beta$  glucuronidase and its substrate  $\beta$ -glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding DME is inserted within a marker gene sequence, transformed cells containing sequences encoding DME can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding DME under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding DME and that express DME may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of DME using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on DME is preferred, but a competitive

binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding DME include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding DME, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding DME may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode DME may be designed to contain signal sequences which direct secretion of DME through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid

sequences encoding DME may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric DME protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of DME activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the DME encoding sequence and the heterologous protein sequence, so that DME may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled DME may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, <sup>35</sup>S-methionine.

DME of the present invention or fragments thereof may be used to screen for compounds that specifically bind to DME. At least one and up to a plurality of test compounds may be screened for specific binding to DME. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of DME, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which DME binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express DME, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E.

coli. Cells expressing DME or cell membrane fractions which contain DME are then contacted with a test compound and binding, stimulation, or inhibition of activity of either DME or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with DME, either in solution or affixed to a solid support, and detecting the binding of DME to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor.

Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

DME of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of DME. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for DME activity, wherein DME is combined with at least one test compound, and the activity of DME in the presence of a test compound is compared with the activity of DME in the absence of the test compound. A change in the activity of DME in the presence of the test compound is indicative of a compound that modulates the activity of DME. Alternatively, a test compound is combined with an in vitro or cell-free system comprising DME under conditions suitable for DME activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of DME may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding DME or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent Number 5,175,383 and U.S. Patent Number 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such

as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

5 Polynucleotides encoding DME may also be manipulated *in vitro* in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) *Science* 282:1145-1147).

10 Polynucleotides encoding DME can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding DME is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with  
15 potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress DME, e.g., by secreting DME in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) *Biotechnol. Annu. Rev.* 4:55-74).

### THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between  
20 regions of DME and drug metabolizing enzymes. In addition, the expression of DME is closely associated with a variety of diseased tissues, including that of the brain, prostate, bone, intestine, and breast. Therefore, DME appears to play a role in autoimmune/inflammatory, cell proliferative, developmental, endocrine, eye, metabolic, and gastrointestinal disorders, including liver disorders. In the treatment of disorders associated with increased DME expression or activity, it is desirable to  
25 decrease the expression or activity of DME. In the treatment of disorders associated with decreased DME expression or activity, it is desirable to increase the expression or activity of DME.

Therefore, in one embodiment, DME or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of DME. Examples of such disorders include, but are not limited to, an autoimmune/inflammatory disorder, such  
30 as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis,

- dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis,
- 5 osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a cell proliferative disorder, such as actinic keratosis, arteriosclerosis,
- 10 atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary,
- 15 pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a developmental disorder, such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary
- 20 neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; an endocrine disorder, such as disorders of the hypothalamus and pituitary resulting from lesions such as primary brain tumors, adenomas, infarction associated with pregnancy, hypophysectomy, aneurysms,
- 25 vascular malformations, thrombosis, infections, immunological disorders, and complications due to head trauma; disorders associated with hypopituitarism including hypogonadism, Sheehan syndrome, diabetes insipidus, Kallman's disease, Hand-Schuller-Christian disease, Letterer-Siwe disease, sarcoidosis, empty sella syndrome, and dwarfism; disorders associated with hyperpituitarism including acromegaly, gigantism, and syndrome of inappropriate antidiuretic hormone (ADH) secretion (SIADH)
- 30 often caused by benign adenoma; disorders associated with hypothyroidism including goiter, myxedema, acute thyroiditis associated with bacterial infection, subacute thyroiditis associated with viral infection, autoimmune thyroiditis (Hashimoto's disease), and cretinism; disorders associated with hyperthyroidism including thyrotoxicosis and its various forms, Grave's disease, pretibial myxedema,

toxic multinodular goiter, thyroid carcinoma, and Plummer's disease; disorders associated with hyperparathyroidism including Conn disease (chronic hypercalcemia); pancreatic disorders such as Type I or Type II diabetes mellitus and associated complications; disorders associated with the adrenals such as hyperplasia, carcinoma, or adenoma of the adrenal cortex, hypertension associated with alkalosis, amyloidosis, hypokalemia, Cushing's disease, Liddle's syndrome, and Arnold-Healy-Gordon syndrome, pheochromocytoma tumors, and Addison's disease; disorders associated with gonadal steroid hormones such as: in women, abnormal prolactin production, infertility, endometriosis, perturbations of the menstrual cycle, polycystic ovarian disease, hyperprolactinemia, isolated gonadotropin deficiency, amenorrhea, galactorrhea, hermaphroditism, hirsutism and virilization, breast cancer, and, in post-menopausal women, osteoporosis; and, in men, Leydig cell deficiency, male climacteric phase, and germinal cell aplasia, hypergonadal disorders associated with Leydig cell tumors, androgen resistance associated with absence of androgen receptors, syndrome of 5  $\alpha$ -reductase, and gynecomastia; an eye disorder, such as conjunctivitis, keratoconjunctivitis sicca, keratitis, episcleritis, iritis, posterior uveitis, glaucoma, amaurosis fugax, ischemic optic neuropathy, optic neuritis, Leber's hereditary optic neuropathy, toxic optic neuropathy, vitreous detachment, retinal detachment, cataract, macular degeneration, central serous chorioretinopathy, retinitis pigmentosa, melanoma of the choroid, retinobulbar tumor, and chiasmal tumor; a metabolic disorder, such as Addison's disease, cerebrotendinous xanthomatosis, congenital adrenal hyperplasia, coumarin resistance, cystic fibrosis, diabetes, fatty hepatocirrhosis, fructose-1,6-diphosphatase deficiency, galactosemia, goiter, glucagonoma, glycogen storage diseases, hereditary fructose intolerance, hyperadrenalism, hypoadrenalism, hyperparathyroidism, hypoparathyroidism, hypercholesterolemia, hyperthyroidism, hypoglycemia, hypothyroidism, hyperlipidemia, hyperlipemia, lipid myopathies, lipodystrophies, lysosomal storage diseases, Menkes syndrome, occipital horn syndrome, mannosidosis, neuraminidase deficiency, obesity, pentosuria phenylketonuria, pseudovitamin D-deficiency rickets; and a gastrointestinal disorder, such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, hereditary hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired immunodeficiency syndrome (AIDS) enteropathy, jaundice,

hepatic encephalopathy, hepatorenal syndrome, hepatic steatosis, hemochromatosis, Wilson's disease, alpha<sub>1</sub>-antitrypsin deficiency, Reye's syndrome, primary sclerosing cholangitis, liver infarction, portal vein obstruction and thrombosis, centrilobular necrosis, peliosis hepatis, hepatic vein thrombosis, veno-occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of pregnancy, and hepatic tumors including nodular hyperplasias, adenomas, and carcinomas.

In another embodiment, a vector capable of expressing DME or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of DME including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified DME in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of DME including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of DME may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of DME including, but not limited to, those listed above.

In a further embodiment, an antagonist of DME may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of DME. Examples of such disorders include, but are not limited to, those autoimmune/inflammatory, cell proliferative, developmental, endocrine, eye, metabolic, and gastrointestinal disorders, including liver disorders described above. In one aspect, an antibody which specifically binds DME may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express DME.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding DME may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of DME including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of DME may be produced using methods which are generally known in the art.

In particular, purified DME may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind DME. Antibodies to DME may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with DME or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to DME have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of DME amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to DME may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies" such as the

D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing *in vivo* production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for DME may also be generated. For example, such fragments include, but are not limited to,  $F(ab)_2$  fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the  $F(ab)_2$  fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between DME and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering DME epitopes is generally used, but a competitive binding assay may also be employed (Pound, *supra*).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for DME. Affinity is expressed as an association constant,  $K_a$ , which is defined as the molar concentration of DME-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The  $K_a$  determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple DME epitopes, represents the average affinity, or avidity, of the antibodies for DME. The  $K_a$  determined for a preparation of monoclonal antibodies, which are monospecific for a particular DME epitope, represents a true measure of affinity. High-affinity antibody preparations with  $K_a$  ranging from about  $10^9$  to  $10^{12}$  L/mole are preferred for use in immunoassays in which the DME-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with  $K_a$  ranging from about  $10^6$  to  $10^7$  L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of DME, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine

the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of DME-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for  
5 antibody quality and usage in various applications, are generally available. (See, e.g., Catty, *supra*, and Coligan et al. *supra*.)

In another embodiment of the invention, the polynucleotides encoding DME, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA,  
10 RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding DME. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding DME. (See, e.g., Agrawal, S., ed. (1996) *Antisense Therapeutics*, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense  
15 sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) *J. Allergy Clin. Immunol.* 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral  
20 vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) *Blood* 76:271; Ausubel, *supra*; Uckert, W. and W. Walther (1994) *Pharmacol. Ther.* 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) *Br. Med. Bull.* 51(1):217-225; Boado, R.J. et al. (1998) *J. Pharm. Sci.* 87(11):1308-1315; and Morris, M.C. et al. (1997) *Nucleic Acids Res.*  
25 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding DME may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) *Science* 288:669-672), severe combined  
30 immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Bläse, R.M. et al. (1995) *Science* 270:475-480; Bordignon, C. et al. (1995) *Science* 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) *Cell* 75:207-216; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:643-666; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:667-703), thalassamias, familial

hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) *Science* 270:404-410; Verma, I.M. and N. Somia (1997) *Nature* 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) *Nature* 335:395-396; Poeschla, E. et al (1996) *Proc. Natl. Acad. Sci. USA* 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in DME expression or regulation causes disease, the expression of DME from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in DME are treated by constructing mammalian expression vectors encoding DME and introducing these vectors by mechanical means into DME-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) *Annu. Rev. Biochem.* 62:191-217; Ivics, Z. (1997) *Cell* 91:501-510; Boulay, J.-L. and H. R  c  pon (1998) *Curr. Opin. Biotechnol.* 9:445-450).

Expression vectors that may be effective for the expression of DME include, but are not limited to, the pCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). DME may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or  $\beta$ -actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) *Proc. Natl. Acad. Sci. USA* 89:5547-5551; Gossen, M. et al. (1995) *Science* 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) *Curr. Opin. Biotechnol.* 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and Blau, H.M. supra), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding DME from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID

TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) *Virology* 52:456-467), or by electroporation (Neumann, E. et al. (1982) *EMBO J.* 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to DME expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding DME under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) *Proc. Natl. Acad. Sci. USA* 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) *J. Virol.* 61:1647-1650; Bender, M.A. et al. (1987) *J. Virol.* 61:1639-1646; Adam, M.A. and A.D. Miller (1988) *J. Virol.* 62:3802-3806; Dull, T. et al. (1998) *J. Virol.* 72:8463-8471; Zufferey, R. et al. (1998) *J. Virol.* 72:9873-9880). U.S. Patent Number 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4<sup>+</sup> T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) *J. Virol.* 71:7020-7029; Bauer, G. et al. (1997) *Blood* 89:2259-2267; Bonyhadi, M.L. (1997) *J. Virol.* 71:4707-4716; Ranga, U. et al. (1998) *Proc. Natl. Acad. Sci. USA* 95:1201-1206; Su, L. (1997) *Blood* 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding DME to cells which have one or more genetic abnormalities with respect to the expression of DME. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csöte, M.E. et al. (1995) *Transplantation* 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent Number 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"),

hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) *Annu. Rev. Nutr.* 19:511-544 and Verma, I.M. and N. Somia (1997) *Nature* 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver  
5 polynucleotides encoding DME to target cells which have one or more genetic abnormalities with respect to the expression of DME. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing DME to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has  
10 been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) *Exp. Eye Res.* 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent Number 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent Number 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a  
15 cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) *J. Virol.* 73:519-532 and Xu, H. et al. (1994) *Dev. Biol.* 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of  
20 multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding DME to target cells. The biology of the prototypic alphavirus, Semliki  
25 Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) *Curr. Opin. Biotechnol.* 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease  
30 and polymerase). Similarly, inserting the coding sequence for DME into the alphavirus genome in place of the capsid-coding region results in the production of a large number of DME-coding RNAs and the synthesis of high levels of DME in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in

hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) *Virology* 228:74-83). The wide host range of alphaviruses will allow the introduction of DME into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding DME.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding DME. Such DNA sequences may be incorporated into a wide variety of vectors

with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible  
5 modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine,  
10 guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding DME. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides,  
15 transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased DME expression or activity, a compound which specifically inhibits expression of the polynucleotide  
20 encoding DME may be therapeutically useful, and in the treatment of disorders associated with decreased DME expression or activity, a compound which specifically promotes expression of the polynucleotide encoding DME may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method  
25 commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a  
30 polynucleotide encoding DME is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an *in vitro* cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding DME are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected

by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding DME. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a *Schizosaccharomyces pombe* gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13).

10 A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruce, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruce, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use *in vivo*, *in vitro*, and *ex vivo*. For *ex vivo* therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

20 Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of *Remington's Pharmaceutical Sciences* (Maack Publishing, Easton PA). Such compositions may consist of DME, antibodies to DME, and mimetics, agonists, antagonists, or inhibitors of DME.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form.

These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising DME or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, DME or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example DME or fragments thereof, antibodies of DME, and agonists, antagonists or inhibitors of DME, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the  $ED_{50}$  (the dose therapeutically effective in 50% of the population) or  $LD_{50}$  (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the  $LD_{50}/ED_{50}$  ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the  $ED_{50}$  with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the

patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1  $\mu\text{g}$  to 100,000  $\mu\text{g}$ , up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

## DIAGNOSTICS

In another embodiment, antibodies which specifically bind DME may be used for the diagnosis of disorders characterized by expression of DME, or in assays to monitor patients being treated with DME or agonists, antagonists, or inhibitors of DME. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for DME include methods which utilize the antibody and a label to detect DME in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring DME, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of DME expression. Normal or standard values for DME expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to DME under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of DME expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding DME may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences,

complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of DME may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of DME, and to monitor regulation of DME levels during therapeutic intervention.

- 5 In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding DME or closely related molecules may be used to identify nucleic acid sequences which encode DME. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding DME, allelic variants, or related sequences.

- Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the DME encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:20-38 or from 15 genomic sequences including promoters, enhancers, and introns of the DME gene.

- Means for producing specific hybridization probes for DNAs encoding DME include the cloning of polynucleotide sequences encoding DME or DME derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA 20 polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as  $^{32}\text{P}$  or  $^{35}\text{S}$ , or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

- Polynucleotide sequences encoding DME may be used for the diagnosis of disorders associated with expression of DME. Examples of such disorders include, but are not limited to, an 25 autoimmune/inflammatory disorder, such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic 30 lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hyper eosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's

syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic  
 lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner  
 syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial,  
 fungal, parasitic, protozoal, and helminthic infections, and trauma; a cell proliferative disorder, such as  
 5 actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue  
 disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis,  
 primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma,  
 myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone,  
 bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung,  
 10 muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus,  
 thyroid, and uterus; a developmental disorder, such as renal tubular acidosis, anemia, Cushing's  
 syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal  
 dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental  
 retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia,  
 15 hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and  
 neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and  
 cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and  
 sensorineural hearing loss; an endocrine disorder, such as disorders of the hypothalamus and pituitary  
 resulting from lesions such as primary brain tumors, adenomas, infarction associated with pregnancy,  
 20 hypophysectomy, aneurysms, vascular malformations, thrombosis, infections, immunological disorders,  
 and complications due to head trauma; disorders associated with hypopituitarism including  
 hypogonadism, Sheehan syndrome, diabetes insipidus, Kallman's disease, Hand-Schüller-Christian  
 disease, Letterer-Siwe disease, sarcoidosis, empty sella syndrome, and dwarfism; disorders associated  
 with hyperpituitarism including acromegaly, gigantism, and syndrome of inappropriate antidiuretic  
 25 hormone (ADH) secretion (SIADH) often caused by benign adenoma; disorders associated with  
 hypothyroidism including goiter, myxedema, acute thyroiditis associated with bacterial infection,  
 subacute thyroiditis associated with viral infection, autoimmune thyroiditis (Hashimoto's disease), and  
 cretinism; disorders associated with hyperthyroidism including thyrotoxicosis and its various forms,  
 Grave's disease, pretibial myxedema, toxic multinodular goiter, thyroid carcinoma, and Plummer's  
 30 disease; disorders associated with hyperparathyroidism including Conn disease (chronic  
 hypercalcemia); pancreatic disorders such as Type I or Type II diabetes mellitus and associated  
 complications; disorders associated with the adrenals such as hyperplasia, carcinoma, or adenoma of  
 the adrenal cortex, hypertension associated with alkalosis, amyloidosis, hypokalemia, Cushing's

- disease, Liddle's syndrome, and Arnold-Healy-Gordon syndrome, pheochromocytoma tumors, and Addison's disease; disorders associated with gonadal steroid hormones such as: in women, abnormal prolactin production, infertility, endometriosis, perturbations of the menstrual cycle, polycystic ovarian disease, hyperprolactinemia, isolated gonadotropin deficiency, amenorrhea, galactorrhea,
- 5 hermaphroditism, hirsutism and virilization, breast cancer, and, in post-menopausal women, osteoporosis; and, in men, Leydig cell deficiency, male climacteric phase, and germinal cell aplasia, hypergonadal disorders associated with Leydig cell tumors, androgen resistance associated with absence of androgen receptors, syndrome of 5  $\alpha$ -reductase, and gynecomastia; an eye disorder, such as conjunctivitis, keratoconjunctivitis sicca, keratitis, episcleritis, iritis, posterior uveitis, glaucoma,
- 10 amaurosis fugax, ischemic optic neuropathy, optic neuritis, Leber's hereditary optic neuropathy, toxic optic neuropathy, vitreous detachment, retinal detachment, cataract, macular degeneration, central serous chorioretinopathy, retinitis pigmentosa, melanoma of the choroid, retrobulbar tumor, and chiasmal tumor; a metabolic disorder, such as Addison's disease, cerebrotendinous xanthomatosis, congenital adrenal hyperplasia, coumarin resistance, cystic fibrosis, diabetes, fatty hepatocirrhosis,
- 15 fructose-1,6-diphosphatase deficiency, galactosemia, goiter, glucagonoma, glycogen storage diseases, hereditary fructose intolerance, hyperadrenalism, hypoadrenalism, hyperparathyroidism, hypoparathyroidism, hypercholesterolemia, hyperthyroidism, hypoglycemia, hypothyroidism, hyperlipidemia, hyperlipemia, lipid myopathies, lipodystrophies, lysosomal storage diseases, Menkes syndrome, occipital horn syndrome, mannosidosis, neuraminidase deficiency, obesity, pentosuria
- 20 phenylketonuria, pseudovitamin D-deficiency rickets; and a gastrointestinal disorder, such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma,
- 25 biliary tract disease, hepatitis, hyperbilirubinemia, hereditary hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired immunodeficiency syndrome (AIDS) enteropathy, jaundice, hepatic encephalopathy, hepatorenal
- 30 syndrome, hepatic steatosis, hemochromatosis, Wilson's disease, alpha<sub>1</sub>-antitrypsin deficiency, Reye's syndrome, primary sclerosing cholangitis, liver infarction, portal vein obstruction and thrombosis, centrilobular necrosis, peliosis hepatis, hepatic vein thrombosis, veno-occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of pregnancy, and hepatic tumors

including nodular hyperplasias, adenomas, and carcinomas. The polynucleotide sequences encoding DME may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered DME expression. Such qualitative or quantitative

5 methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding DME may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding DME may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable  
10 incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding DME in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor  
15 the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of DME, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding DME, under conditions suitable for hybridization or amplification.

20 Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

25 Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

30 With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ

preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding DME may involve the use of PCR. These oligomers may be chemically synthesized, generated  
5 enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding DME, or a fragment of a polynucleotide complementary to the polynucleotide encoding DME, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

10 In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding DME may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers  
15 derived from the polynucleotide sequences encoding DME are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSSCP, the oligonucleotide primers are  
20 fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed *in silico* SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing  
25 errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

Methods which may also be used to quantify the expression of DME include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from  
30 standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid

quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, DME, fragments of DME, or antibodies specific for DME may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent Number 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental

compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) *Mol. Carcinog.* 24:153-159; Steiner, S. and N.L. Anderson (2000) *Toxicol. Lett.* 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at <http://www.niehs.nih.gov/oc/news/toxchip.htm>.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl

sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, *supra*). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently  
5 positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial  
10 sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for DME to quantify the levels of DME expression. In one embodiment, the antibodies are used as elements on a microarray,  
15 and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) *Anal. Biochem.* 270:103-111; Mendoz, L.G. et al. (1999) *Biotechniques* 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array  
20 element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) *Electrophoresis* 18:533-537), so proteome toxicant signatures may be  
25 useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated  
30 biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid

residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding DME may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent *in situ* hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, *supra*, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding DME on a

physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps.

- 5 Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any  
10 sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

- In another embodiment of the invention, DME, its catalytic or immunogenic fragments, or  
15 oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between DME and the agent being tested may be measured.

- Another technique for drug screening provides for high throughput screening of compounds  
20 having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with DME, or fragments thereof, and washed. Bound DME is then detected by methods well known in the art. Purified DME can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively,  
25 non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding DME specifically compete with a test compound for binding DME. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with DME.

- 30 In additional embodiments, the nucleotide sequences which encode DME may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

5 The disclosures of all patents, applications and publications, mentioned above and below, including U.S. Ser. No. 60/223,055, U.S. Ser. No. 60/224,728, U.S. Ser. No. 60/226,440, U.S. Ser. No. 60/228,067, U.S. Ser. No. 60/230,063, U.S. Ser. No. 60/232,244, and U.S. Ser. No. 60/234,269, are expressly incorporated by reference herein.

## 10 EXAMPLES

### I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA) and shown in Table 4, column 5. Some tissues were  
15 homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

20 Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)<sup>+</sup> RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the  
25 POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERScript plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, *supra*,  
30 units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column

chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA (Invitrogen), or pINCY (Incyte Genomics, Palo Alto CA), or derivatives thereof. Recombinant plasmids were transformed into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 $\alpha$ , DH10B, or ElectroMAX DH10B from Life Technologies.

## II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by *in vivo* excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PRBP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

## III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading

frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing  
5 vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and hidden Markov model (HMM)-based protein family  
10 databases such as PFAM. (HMM is a probabilistic approach which analyzes consensus primary structures of gene families. See, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or  
15 Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may  
20 begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and hidden Markov model (HMM)-based protein family databases such as PFAM. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering,  
25 South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of  
30 Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where

applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:20-38. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 4.

#### IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative drug metabolizing enzymes were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpr and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode drug metabolizing enzymes, the encoded polypeptides were analyzed by querying against PFAM models for drug metabolizing enzymes. Potential drug metabolizing enzymes were also identified by homology to Incyte cDNA sequences that had been annotated as drug metabolizing enzymes. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpr public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

#### V. Assembly of Genomic Sequence Data with cDNA Sequence Data

##### "Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene

identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbprl public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

#### "Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

#### **VI. Chromosomal Mapping of DME Encoding Polynucleotides**

The sequences which were used to assemble SEQ ID NO:20-38 were compared with

sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:20-38 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO., to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (<http://www.ncbi.nlm.nih.gov/genemap/>), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

## VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, *supra*, ch. 7; Ausubel (1995) *supra*, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum \{length(Seq. 1), length(Seq. 2)\}}}$$

The product score takes into account both the degree of similarity between two sequences and the

length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding DMB are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding DMB. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

#### VIII. Extension of DME Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30

nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one  
5 extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing  $Mg^{2+}$ ,  $(NH_4)_2SO_4$ , and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE  
10 enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4  
15 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100  $\mu$ l PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5  $\mu$ l of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II  
20 (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5  $\mu$ l to 10  $\mu$ l aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and  
25 sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector  
30 (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was  
5 quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethylsulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

10 In like manner, full length polynucleotide sequences are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

#### IX. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:20-38 are employed to screen cDNAs,  
15 genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250  $\mu$ Ci of [ $\gamma$ -<sup>32</sup>P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase  
20 (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10<sup>7</sup> counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

25 The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and  
30 compared.

#### X. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, *supra.*), mechanical

microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schna (1999), *supra*).

Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface

5 of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schna, M. et al. (1995) *Science* 270:467-470; Shalon, D. et al. (1996) *Genome Res.* 6:639-645; Marshall, A. and J. Hodgson (1998) *Nat. Biotechnol.* 16:27-31.)

10 Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection.

15 After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorption and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described  
20 in detail below.

#### Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)<sup>+</sup> RNA is purified using the oligo-(dT) cellulose method. Each poly(A)<sup>+</sup> RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/ $\mu$ l oligo-(dT) primer (21mer), 1X first  
25 strand buffer, 0.03 units/ $\mu$ l RNase inhibitor, 500  $\mu$ M dATP, 500  $\mu$ M dGTP, 500  $\mu$ M dTTP, 40  $\mu$ M dCTP, 40  $\mu$ M dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)<sup>+</sup> RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)<sup>+</sup> RNAs are synthesized by *in vitro* transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one  
30 with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated

using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14  $\mu$ l 5X SSC/0.2% SDS.

#### Microarray Preparation

5 Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5  $\mu$ g. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

10 Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C  
15 oven.

Array elements are applied to the coated glass substrate using a procedure described in US Patent No. 5,807,522, incorporated herein by reference. 1  $\mu$ l of the array element DNA, at an average concentration of 100 ng/ $\mu$ l, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

20 Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

#### 25 Hybridization

Hybridization reactions contain 9  $\mu$ l of sample mixture consisting of 0.2  $\mu$ g each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm<sup>2</sup> coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly  
30 larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140  $\mu$ l of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

### Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used

for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

#### XI. Complementary Polynucleotides

Sequences complementary to the DME-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring DME. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of DME. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the DME-encoding transcript.

#### XII. Expression of DME

Expression and purification of DME is achieved using bacterial or virus-based expression systems. For expression of DME in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac (tac)* hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express DME upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of DME in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding DME by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, DME is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from DME at

specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, *supra*, ch. 10 and 16). Purified DME obtained by these methods can be used directly in the assays shown in Examples XVI, XVII, and XVIII, where applicable.

### XIII. Functional Assays

DME function is assessed by expressing the sequences encoding DME at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10  $\mu$ g of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2  $\mu$ g of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of DME on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding DME and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success

NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding DME and other genes of interest can be analyzed by northern analysis or microarray techniques.

#### XIV. Production of DME Specific Antibodies

5 DME substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the DME amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is  
10 synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, *supra*, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-  
15 Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, *supra*.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-DME activity by, for example, binding the peptide or DME to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit  
20 IgG.

#### XV. Purification of Naturally Occurring DME Using Specific Antibodies

Naturally occurring or recombinant DME is substantially purified by immunoaffinity chromatography using antibodies specific for DME. An immunoaffinity column is constructed by covalently coupling anti-DME antibody to an activated chromatographic resin, such as CNBr-activated  
25 SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing DME are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of DME (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/DME binding  
30 (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and DME is collected.

#### XVI. Identification of Molecules Which Interact with DME

DME, or biologically active fragments thereof, are labeled with  $^{125}\text{I}$  Bolton-Hunter reagent. (See, e.g., Bolton A.B. and W.M. Hunter (1973) *Biochem. J.* 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled DME, washed, and any wells with labeled DME complex are assayed. Data obtained using different concentrations of DME are used to calculate values for the number, affinity, and association of DME with the candidate molecules.

Alternatively, molecules interacting with DME are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989) *Nature* 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

DME may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

#### XVII. Demonstration of DME Activity

Cytochrome P450 activity of DME is measured using the 4-hydroxylation of aniline. Aniline is converted to 4-aminophenol by the enzyme, and has an absorption maximum at 630 nm (Gibson and Skett, *supra*). This assay is a convenient measure, but underestimates the total hydroxylation, which also occurs at the 2- and 3- positions. Assays are performed at 37°C and contain an aliquot of the enzyme and a suitable amount of aniline (approximately 2 mM) in reaction buffer. For this reaction, the buffer must contain NADPH or an NADPH-generating cofactor system. One formulation for this reaction buffer includes 85 mM Tris pH 7.4, 15 mM  $\text{MgCl}_2$ , 50 mM nicotinamide, 40 mg trisodium isocitrate, and 2 units isocitrate dehydrogenase, with 8 mg  $\text{NADP}^+$  added to a 10 mL reaction buffer stock just prior to assay. Reactions are carried out in an optical cuvette, and the absorbance at 630 nm is measured. The rate of increase in absorbance is proportional to the enzyme activity in the assay. A standard curve can be constructed using known concentrations of 4-aminophenol.

$1\alpha,25$ -dihydroxyvitamin D 24-hydroxylase activity of DME is determined by monitoring the conversion of  $^3\text{H}$ -labeled  $1\alpha,25$ -dihydroxyvitamin D ( $1\alpha,25(\text{OH})_2\text{D}$ ) to  $24,25$ -dihydroxyvitamin D ( $24,25(\text{OH})_2\text{D}$ ) in transgenic rats expressing DME. 1  $\mu\text{g}$  of  $1\alpha,25(\text{OH})_2\text{D}$  dissolved in ethanol (or ethanol alone as a control) is administered intravenously to approximately 6-week-old male transgenic rats expressing DME or otherwise identical control rats expressing either a defective variant of DME or not expressing DME. The rats are killed by decapitation after 8 hrs, and the kidneys are rapidly removed, rinsed, and homogenized in 9 volumes of ice-cold buffer (15 mM Tris-acetate (pH 7.4), 0.19 M sucrose, 2 mM magnesium acetate, and 5 mM sodium succinate). A portion (e.g., 3 ml) of each

homogenate is then incubated with 0.25 nM  $1\alpha,25(\text{OH})_2[1\text{-}^3\text{H}]\text{D}$ , with a specific activity of approximately 3.5 GBq/mmol, for 15 min at 37 °C under oxygen with constant shaking. Total lipids are extracted as described (Bligh, E.G. and W.J. Dyer (1959) Can. J. Biochem. Physiol. 37: 911-917) and the chloroform phase is analyzed by HPLC using a FINEPAK SIL column (JASCO, Tokyo, Japan) with a *n*-hexane/chloroform/methanol (10:2.5:1.5) solvent system at a flow rate of 1 ml/min. In the alternative, the chloroform phase is analyzed by reverse phase HPLC using a J SPHERE ODS-AM column (YMC Co. Ltd., Kyoto, Japan) with an acetonitrile buffer system (40 to 100%, in water, in 30 min) at a flow rate of 1 ml/min. The eluates are collected in fractions of 30 seconds (or less) and the amount of  $^3\text{H}$  present in each fraction is measured using a scintillation counter. By comparing the chromatograms of control samples (i.e., samples comprising  $1\alpha,25$ -dihydroxyvitamin D or  $24,25$ -dihydroxyvitamin D ( $24,25(\text{OH})_2\text{D}$ ), with the chromatograms of the reaction products, the relative mobilities of the substrate ( $1\alpha,25(\text{OH})_2[1\text{-}^3\text{H}]\text{D}$ ) and product ( $24,25(\text{OH})_2[1\text{-}^3\text{H}]\text{D}$ ) are determined and correlated with the fractions collected. The amount of  $24,25(\text{OH})_2[1\text{-}^3\text{H}]\text{D}$  produced in control rats is subtracted from that of transgenic rats expressing DME. The difference in the production of  $24,25(\text{OH})_2[1\text{-}^3\text{H}]\text{D}$  in the transgenic and control animals is proportional to the amount of 25-hydroxylase activity of DME present in the sample. Confirmation of the identity of the substrate and product(s) is confirmed by means of mass spectroscopy (Miyamoto, Y. et al. (1997) J. Biol. Chem. 272:14115-14119).

Flavin-containing monooxygenase activity of DME is measured by chromatographic analysis of metabolic products. For example, Ring, B. J. et al. (1999; Drug Metab. Dis. 27:1099-1103) incubated FMO in 0.1 M sodium phosphate buffer (pH 7.4 or 8.3) and 1 mM NADPH at 37 °C, stopped the reaction with an organic solvent, and determined product formation by HPLC. Alternatively, activity is measured by monitoring oxygen uptake using a Clark-type electrode. For example, Ziegler, D. M. and Poulsen, L. L. (1978; Methods Enzymol. 52:142-151) incubated the enzyme at 37 °C in an NADPH-generating cofactor system (similar to the one described above) containing the substrate methimazole. The rate of oxygen uptake is proportional to enzyme activity.

UDP glucuronyltransferase activity of DME is measured using a colorimetric determination of free amine groups (Gibson and Skett, *supra*). An amine-containing substrate, such as 2-aminophenol, is incubated at 37 °C with an aliquot of the enzyme in a reaction buffer containing the necessary cofactors (40 mM Tris pH 8.0, 7.5 mM  $\text{MgCl}_2$ , 0.025% Triton X-100, 1 mM ascorbic acid, 0.75 mM UDP-glucuronic acid). After sufficient time, the reaction is stopped by addition of ice-cold 20% trichloroacetic acid in 0.1 M phosphate buffer pH 2.7, incubated on ice, and centrifuged to clarify the supernatant. Any unreacted 2-aminophenol is destroyed in this step. Sufficient freshly-prepared

sodium nitrite is then added; this step allows formation of the diazonium salt of the glucuronidated product. Excess nitrite is removed by addition of sufficient ammonium sulfamate, and the diazonium salt is reacted with an aromatic amine (for example, N-naphthylethylenediamine) to produce a colored azo compound which can be assayed spectrophotometrically (at 540 nm for the example). A standard curve can be constructed using known concentrations of aniline, which will form a chromophore with similar properties to 2-aminophenol glucuronide.

Sulfotransferase activity of DME is measured using the incorporation of  $^{35}\text{S}$  from [ $^{35}\text{S}$ ]PAPS into a model substrate such as phenol (Folds, A. and Meek, J. L. (1973) *Biochim. Biophys. Acta* 327:365-374). An aliquot of enzyme is incubated at 37°C with 1 mL of 10 mM phosphate buffer pH 6.4, 50  $\mu\text{M}$  phenol, 0.4-4.0  $\mu\text{M}$  [ $^{35}\text{S}$ ]PAPS. After sufficient time for 5-20% of the radiolabel to be transferred to the substrate, 0.2 mL of 0.1 M barium acetate is added to precipitate protein and phosphate buffer. Then 0.2 mL of 0.1 M  $\text{Ba}(\text{OH})_2$  is added, followed by 0.2 mL  $\text{ZnSO}_4$ . The supernatant is cleared by centrifugation, which removes proteins as well as unreacted [ $^{35}\text{S}$ ]PAPS. Radioactivity in the supernatant is measured by scintillation. The enzyme activity is determined from the number of moles of radioactivity in the reaction product.

Glutathione S-transferase activity of DME is measured using a model substrate, such as 2,4-dinitro-1-chlorobenzene, which reacts with glutathione to form a product, 2,4-dinitrophenyl-glutathione, that has an absorbance maximum at 340 nm. It is important to note that GSTs have differing substrate specificities, and the model substrate should be selected based on the substrate preferences of the GST of interest. Assays are performed at ambient temperature and contain an aliquot of the enzyme in a suitable reaction buffer (for example, 1 mM glutathione, 1 mM dinitrochlorobenzene, 90 mM potassium phosphate buffer pH 6.5). Reactions are carried out in an optical cuvette, and the absorbance at 340 nm is measured. The rate of increase in absorbance is proportional to the enzyme activity in the assay.

N-acyltransferase activity of DME is measured using radiolabeled amino acid substrates and measuring radiolabel incorporation into conjugated products. Enzyme is incubated in a reaction buffer containing an unlabeled acyl-CoA compound and radiolabeled amino acid, and the radiolabeled acyl-conjugates are separated from the unreacted amino acid by extraction into n-butanol or other appropriate organic solvent. For example, Johnson, M. R. et al. (1990; *J. Biol. Chem.* 266:10227-10233) measured bile acid-CoA:amino acid N-acyltransferase activity by incubating the enzyme with cholyl-CoA and  $^3\text{H}$ -glycine or  $^3\text{H}$ -taurine, separating the tritiated cholate conjugate by extraction into n-butanol, and measuring the radioactivity in the extracted product by scintillation. Alternatively, N-acyltransferase activity is measured using the spectrophotometric determination of reduced CoA

(CoASH) described below.

N-acetyltransferase activity of DME is measured using the transfer of radiolabel from [ $^{14}\text{C}$ ]acetyl-CoA to a substrate molecule (for example, see Deguchi, T. (1975) *J. Neurochem.* 24:1083-5). Alternatively, a newer spectrophotometric assay based on DTNB (5,5'-dithio-bis(2-nitrobenzoic acid; Ellman's reagent) reaction with CoASH may be used. Free thiol-containing CoASH is formed during N-acetyltransferase catalyzed transfer of an acetyl group to a substrate. CoASH is detected using the absorbance of DTNB conjugate at 412 nm (De Angelis, J. et al. (1997) *J. Biol. Chem.* 273:3045-3050). Enzyme activity is proportional to the rate of radioactivity incorporation into substrate, or the rate of absorbance increase in the spectrophotometric assay.

Protein arginine methyltransferase activity of DME is measured at 37 °C for various periods of time. S-adenosyl-L-[methyl- $^3\text{H}$ ]methionine ( $^3\text{H}$ ]AdoMet; specific activity = 75 Ci/mmol; NEN Life Science Products) is used as the methyl-donor substrate. Useful methyl-accepting substrates include glutathione S-transferase fibrillarin glycine-arginine domain fusion protein (GST-GAR), heterogeneous nuclear ribonucleoprotein (hnRNP), or hypomethylated proteins present in lysates from adenosine dialdehyde-treated cells. Methylation reactions are stopped by adding SDS-PAGE sample buffer. The products of the reactions are resolved by SDS-PAGE and visualized by fluorography. The presence of  $^3\text{H}$ -labeled methyl-donor substrates is indicative of protein arginine methyltransferase activity of DME (Tang, J. et al. (2000) *J. Biol. Chem.* 275:7723-7730 and Tang, J. et al. (2000) *J. Biol. Chem.* 275:19866-19876).

Aldo/keto reductase activity of DME is measured using the decrease in absorbance at 340 nm as NADPH is consumed. A standard reaction mixture is 135 mM sodium phosphate buffer (pH 6.2-7.2 depending on enzyme), 0.2 mM NADPH, 0.3 M lithium sulfate, 0.5-2.5  $\mu\text{g}$  enzyme and an appropriate level of substrate. The reaction is incubated at 30°C and the reaction is monitored continuously with a spectrophotometer. Enzyme activity is calculated as mol NADPH consumed /  $\mu\text{g}$  of enzyme.

Alcohol dehydrogenase activity of DME is measured using the increase in absorbance at 340 nm as  $\text{NAD}^+$  is reduced to NADH. A standard reaction mixture is 50 mM sodium phosphate, pH 7.5, and 0.25 mM EDTA. The reaction is incubated at 25°C and monitored using a spectrophotometer. Enzyme activity is calculated as mol NADH produced /  $\mu\text{g}$  of enzyme.

DME activity is determined using 4-methylumbelliferyl acetate as a substrate. The enzymatic reaction is initiated by adding approximately 10  $\mu\text{l}$  of DME-containing sample to 1 ml of reaction buffer (90 mM  $\text{KH}_2\text{PO}_4$ , 40 mM KCl, pH 7.3) with 0.5 mM 4-methylumbelliferyl acetate at 37 °C. The production of 4-methylumbelliferone is monitored with a spectrophotometer ( $\epsilon_{350} =$

12.2 mM<sup>-1</sup> cm<sup>-1</sup>) for 1.5 min. Specific activity is expressed as micromoles of product formed per minute per milligram of protein and corresponds to the activity of DMB in the sample (Evgenia, V. et al. (1997) J. Biol. Chem. 272:14769-14775).

In the alternative, the cocaine benzoyl ester hydrolase activity of DMB is measured by incubating approximately 0.1 ml of enzyme 3.3 mM cocaine in reaction buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, pH 7.4) with 1 mM benzamidine, 1 mM EDTA, and 1 mM dithiothreitol at 37 °C. The reaction is incubated for 1 h in a total volume of 0.4 ml then terminated with an equal volume of 5% trichloroacetic acid. 0.1 ml of the internal standard 3,4-dimethylbenzoic acid (10 µg/ml) is added. Precipitated protein is separated by centrifugation at 12,000 × g for 10 min. The supernatant is transferred to a clean tube and extracted twice with 0.4 ml of methylene chloride. The two extracts are combined and dried under a stream of nitrogen. The residue is resuspended in 14% acetonitrile, 250 mM KH<sub>2</sub>PO<sub>4</sub>, pH 4.0, with 8 µl of diethylamine per 100 ml and ) and injected onto a C18 reverse-phase HPLC column for separation. The column eluate was monitored at 235 nm. DMB activity is quantified by comparing peak area ratios of the analyte to the internal standard. A standard curve was generated with benzoic acid standards prepared in a trichloroacetic acid-treated protein matrix (Evgenia, V. et al. (1997) J. Biol. Chem. 272:14769-14775).

In another alternative, DMB carboxyl esterase activity against the water-soluble substrate para-nitrophenyl butyric acid is determined by spectrophotometric methods well known to those skilled in the art. In this procedure, the DMB-containing samples are diluted with 0.5 M Tris-HCl (pH 7.4 or 8.0) or sodium acetate (pH 5.0) in the presence of 6 mM taurocholate. The assay is initiated by adding a freshly prepared para-nitrophenyl butyric acid solution (100 µg/ml in sodium acetate, pH 5.0). Carboxyl esterase activity was then monitored and compared with control autohydrolysis of the substrate using an spectrophotometer set at 405 nm (Wan, L. et al. (2000) J. Biol. Chem. 275:10041-10046).

Heparan sulfate 6-sulfotransferase activity of DMB is measured in vitro by incubating a sample containing DMB along with 2.5 µmol imidazole HCl (pH 6.8), 3.75 µg of protamine chloride, 25 nmol (as hexosamine) of completely desulfated and N-resulfated heparin, and 50 pmol (about 5 × 10<sup>5</sup> cpm) of [<sup>35</sup>S] adenosine 3'-phosphate 5'-phosphosulfate (PAPS) in a final reaction volume of 50 µl at 37 °C for 20 min. The reaction is stopped by immersing the reaction tubes in a boiling water bath for 1 min. 0.1 µmol (as glucuronic acid) of chondroitin sulfate A is added to the reaction mixture as a carrier. <sup>35</sup>S-labeled polysaccharides are precipitated with 3 volumes of cold ethanol containing 1.3%

potassium acetate and separated completely from unincorporated [ $^{35}\text{S}$ ]PAPS and its degradation products by gel chromatography using desalting columns. One unit of enzyme activity is defined as the amount required to transfer 1 pmol of sulfate/min. as determined by the amount of [ $^{35}\text{S}$ ]PAPS incorporated into the precipitated polysaccharides (Habuchi, H. et al. (1995) J. Biol. Chem.

5 270:4172-4179).

In the alternative, heparan sulfate 6-sulfotransferase activity of DME is measured by extraction and renaturation of enzyme from gels following separation by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Following separation, the gel is washed with 0.05 M Tris-HCl, pH 8.0, cut into 3-5 mm segments and subjected to agitation at 4 °C with 100  $\mu\text{l}$  of 0.05 M Tris-HCl, pH 8.0 containing 0.15 M NaCl for 48 h. The eluted enzyme is collected by centrifugation and assayed for the sulfotransferase activity as above (Habuchi, H. et al. (1995) J. Biol. Chem. 270:4172-4179).

In another alternative, DME sulfotransferase activity is determined by measuring the transfer of [ $^{35}\text{S}$ ]sulfate from [ $^{35}\text{S}$ ]PAPS to an immobilized peptide. In one example, the peptide (QATEYBYLDYDFLPEC) represents the N-terminal 15 residues of the mature P-selectin glycoprotein ligand-1 polypeptide to which is added C-terminal cysteine residue. The peptide spans three potential tyrosine sulfation sites. The peptide is linked via the cysteine residue to iodoacetamide-activated resin at a density of 1.5-3.0  $\mu\text{mol}$  peptide/ml of resin. The enzyme assay is performed by combining 10  $\mu\text{l}$  of peptide-derivitized beads with 2-20  $\mu\text{l}$  of DME-containing sample in 40 mM Pipes (pH 6.8), 0.3 M NaCl, 20 mM  $\text{MnCl}_2$ , 50 mM NaF, 1% Triton X-100, and 1 mM 5'-AMP in a final volume of 130  $\mu\text{l}$ . The assay is initiated by addition of 0.5  $\mu\text{Ci}$  of [ $^{35}\text{S}$ ]PAPS (1.7  $\mu\text{M}$ ; 1 Ci = 37 GBq). After 30 min at 37°C, the reaction beads are washed with 6 M guanidine at 65°C and the radioactivity incorporated into the beads is determined by liquid scintillation counting. Transfer of [ $^{35}\text{S}$ ]sulfate to the bead-associated peptide is measured to determine the DME activity in the sample. One unit of activity is defined as 1 pmol of product formed per min (Ouyang, Y-B. et al. (1998) Biochemistry 95:2896-2901).

In another alternative, DME sulfotransferase assays are performed using [ $^{35}\text{S}$ ]PAPS as the sulfate donor in a final volume of 30  $\mu\text{l}$ , contains 50 mM Hepes-NaOH (pH 7.0), 250 mM sucrose, 1 mM dithiothreitol, 14  $\mu\text{M}$  [ $^{35}\text{S}$ ]PAPS (15 Ci/mmol), and dopamine (25  $\mu\text{M}$ ), *p*-nitrophenol (5  $\mu\text{M}$ ), or other candidate substrates. Assay reactions are started by the addition of a purified DME enzyme preparation or a sample containing DME activity, allowed to proceed for 15 min at 37 °C, and terminated by heating at 100 °C for 3 min. The precipitates formed are cleared by

centrifugation. The supernatants are then subjected to the analysis of  $^{35}\text{S}$ -sulfated product by either thin-layer chromatography or a two-dimensional thin layer separation procedure. Appropriate standards are run in parallel with the supernatants to allow the identification of the  $^{35}\text{S}$ -sulfated products and determine the enzyme specificity of the DME-containing samples based on relative  
5 relates of migration of reaction products (Sakakibara, Y. et al. (1998) J. Biol. Chem. 273:6242-6247).

Squalene epoxidase activity of DME is assayed in a mixture comprising purified DME (or a crude mixture comprising DME), 20 mM Tris-HCl (pH 7.5), 0.01 mM FAD, 0.2 unit of NADPH-cytochrome C (P-450) reductase, 0.01 mM [ $^{14}\text{C}$ ]squalene (dispersed with the aid of 20  $\mu\text{l}$  of  
10 Tween 80), and 0.2% Triton X-100. 1 mM NADPH is added to initiate the reaction followed by incubation at 37 °C for 30 min. The nonsaponifiable lipids are analyzed by silica gel TLC developed with ethyl acetate/benzene (0.5:99.5, v/v). The reaction products are compared to those from a reaction mixture without DME. The presence of 2,3(*S*)-oxidosqualene is confirmed using appropriate lipid standards (Sakakibara, J. et al. (1995) 270:17-20).

15 Epoxide hydrolase activity of DME is determined by following substrate depletion using gas chromatographic (GC) analysis of ethereal extracts or by following substrate depletion and diol production by GC analysis of reaction mixtures quenched in acetone. A sample containing DME or an epoxide hydrolase control sample is incubated in 10 mM Tris-HCl (pH 8.0), 1 mM ethylenediaminetetraacetate (EDTA), and 5 mM epoxide substrate (e.g., ethylene oxide, styrene  
20 oxide, propylene oxide, isoprene monoxide, epichlorohydrin, epibromohydrin, epifluorohydrin, glycidol, 1,2-epoxybutane, 1,2-epoxyhexane, or 1,2-epoxyoctane). A portion of the sample is withdrawn from the reaction mixture at various time points, and added to 1 ml of ice-cold acetone containing an internal standard for GC analysis (e.g., 1-nonanol). Protein and salts are removed by centrifugation (15 min, 4000  $\times$  g) and the extract is analyzed by GC using a 0.2 mm  $\times$  25-m CP-Wax57-CB column  
25 (CHROMPACK, Middelburg, The Netherlands) and a flame-ionization detector. The identification of GC products is performed using appropriate standards and controls well known to those skilled in the art. 1 Unit of DME activity is defined as the amount of enzyme that catalyzes the production of 1  $\mu\text{mol}$  of diol/min (Rink, R. et al. (1997) J. Biol. Chem. 272:14650-14657).

Aminotransferase activity of DME is assayed by incubating samples containing DME for 1  
30 hour at 37 °C in the presence of 1 mM L-kynurenine and 1 mM 2-oxoglutarate in a final volume of 200  $\mu\text{l}$  of 150 mM Tris acetate buffer (pH 8.0) containing 70  $\mu\text{M}$  PLP. The formation of kynurenic acid is quantified by HPLC with spectrophotometric detection at 330 nm using the appropriate standards and controls well known to those skilled in the art. In the alternative,

L-3-hydroxykynurenine is used as substrate and the production of xanthurenic acid is determined by HPLC analysis of the products with UV detection at 340 nm. The production of kynurenic acid xanthurenic acid, respectively, is indicative of aminotransferase activity (Buchli, R. et al. (1995) *J. Biol. Chem.* 270:29330-29335).

5 In another alternative, aminotransferase activity of DMB is measured by determining the activity of purified DMB or crude samples containing DMB toward various amino and oxo acid substrates under single turnover conditions by monitoring the changes in the UV/VIS absorption spectrum of the enzyme-bound cofactor, PLP. The reactions are performed at 25 °C in 50 mM 4-methylmorpholine (pH 7.5) containing 9 µM purified DMB or DMB containing samples and  
10 substrate to be tested (amino and oxo acid substrates). The half-reaction from amino acid to oxo acid is followed by measuring the decrease in absorbance at 360 nm and the increase in absorbance at 330 nm due to the conversion of enzyme-bound PLP to PMP. The specificity and relative activity of DMB is determined by the activity of the enzyme preparation against specific substrates (Vacca, R.A. et al. (1997) *J. Biol. Chem.* 272:21932-21937).

15 Superoxide dismutase activity of DMB is assayed from cell pellets, culture supernatants, or purified protein preparations. Samples or lysates are resolved by electrophoresis on 15% non-denaturing polyacrylamide gels. The gels are incubated for 30 min in 2.5 mM nitro blue tetrazolium, followed by incubation for 20 min in 30 mM potassium phosphate, 30 mM TEMED, and 30 µM riboflavin (pH 7.8). Superoxide dismutase activity is visualized as white bands against a blue  
20 background, following illumination of the gels on a lightbox. Quantitation of superoxide dismutase activity is performed by densitometric scanning of the activity gels using the appropriate superoxide dismutase positive and negative controls (e.g., various amounts of commercially available *E. coli* superoxide dismutase (Harth, G. and Horwitz, M.A. (1999) *J. Biol. Chem.* 274:4281-4292).

Catechol-*O*-methyltransferase activity of DMB is measured in a reaction mixture consisting  
25 of 50 mM Tris-HCl (pH 7.4), 1.2 mM MgCl<sub>2</sub>, 200 µM SAM (*S*-adenosyl-L-methionine) iodide (containing 0.5 µCi of [methyl-<sup>3</sup>H]SAM), 1 mM dithiothreitol, and varying concentrations of catechol substrate (e.g., L-dopa, dopamine, or DBA) in a final volume of 1.0 ml. The reaction is initiated by the addition of 250-500 µg of purified DMB or crude DMB-containing sample and performed at 37 °C for 30 min. The reaction is arrested by rapidly cooling on ice and immediately  
30 extracting with 7 ml of ice-cold n-heptane. Following centrifugation at 1000 x g for 10 min, 3-ml aliquots of the organic extracts are analyzed for radioactivity content by liquid scintillation counting. The level of catechol-associated radioactivity in the organic phase is proportional to the activity

catechol-*O*-methyltransferase activity of DME (Zhu, B.T. Liehr, J.G. (1996) 271:1357-1363).

DHFR activity of DME is determined spectrophotometrically at 15 °C by following the disappearance of NADPH at 340 nm ( $\epsilon_{340} = 11,800 \text{ M}^{-1}\text{cm}^{-1}$ ). The standard assay mixture contains 100  $\mu\text{M}$  NADPH, 14 mM 2-mercaptoethanol, MTEN buffer (50 mM 2-morpholinoethanesulfonic acid, 25 mM tris(hydroxymethyl)aminomethane, 25 mM ethanolamine, and 100 mM NaCl, pH 7.0), and DME in a final volume of 2.0 ml. The reaction is started by the addition of 50  $\mu\text{M}$  dihydrofolate (as substrate). The oxidation of NADPH to  $\text{NADP}^+$  corresponds to the reduction of dihydrofolate in the reaction and is proportional to the amount of DHFR activity in the sample (Nakamura, T. and Iwakura, M. (1999) J. Biol. Chem. 274:19041-19047).

Sulfotransferase activity of DME is measured using the incorporation of  $^{35}\text{S}$  from [ $^{35}\text{S}$ ]PAPS into a model substrate such as phenol (Folds, A. and Meek, J. L. (1973) Biochim. Biophys. Acta 327:365-374). An aliquot of enzyme is incubated at 37°C with 1 mL of 10 mM phosphate buffer pH 6.4, 50  $\mu\text{M}$  phenol, 0.4-4.0  $\mu\text{M}$  [ $^{35}\text{S}$ ]PAPS. After sufficient time for 5-20% of the radiolabel to be transferred to the substrate, 0.2 mL of 0.1 M barium acetate is added to precipitate protein and phosphate buffer. Then 0.2 mL of 0.1 M  $\text{Ba}(\text{OH})_2$  is added, followed by 0.2 mL  $\text{ZnSO}_4$ . The supernatant is cleared by centrifugation, which removes proteins as well as unreacted [ $^{35}\text{S}$ ]PAPS. Radioactivity in the supernatant is measured by scintillation. The enzyme activity is determined from the number of moles of radioactivity in the reaction product.

#### XVIII. Identification of DME Inhibitors

Compounds to be tested are arrayed in the wells of a multi-well plate in varying concentrations along with an appropriate buffer and substrate, as described in the assays in Example XVII. DME activity is measured for each well and the ability of each compound to inhibit DME activity can be determined, as well as the dose-response profiles. This assay could also be used to identify molecules which enhance DME activity.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments.

Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
7248285	1	7248285CD1	20	7248285CB1
7472835	2	7472835CD1	21	7472835CB1
7476203	3	7476203CD1	22	7476203CB1
7478583	4	7478583CD1	23	7478583CB1
7478585	5	7478585CD1	24	7478585CB1
7479904	6	7479904CD1	25	7479904CB1
7480367	7	7480367CD1	26	7480367CB1
8069390	8	8069390CD1	27	8069390CB1
7473869	9	7473869CD1	28	7473869CB1
7478588	10	7478588CD1	29	7478588CB1
55046125	11	55046125CD1	30	55046125CB1
3538709	12	3538709CD1	31	3538709CB1
71563101	13	71563101CD1	32	71563101CB1
7472027	14	7472027CD1	33	7472027CB1
7480358	15	7480358CD1	34	7480358CB1
1618256	16	1618256CD1	35	1618256CB1
3387823	17	3387823CD1	36	3387823CB1
55142051	18	55142051CD1	37	55142051CB1
7395274	19	7395274CD1	38	7395274CB1

Table 2

Polyptide SEQ ID NO:	Incyte Polyptide ID	GenBank ID NO:	Probability score	GenBank Homolog
1	7248285CD1	g4929199	1.90E-248	[Perca flavescens] lysyl oxidase related protein homolog Langenau, D.M. et al. (1999) J. Mol. Endocrinol. 23:137-152.
2	7472835CD1	g559046	4.50E-181	[Homo sapiens] flavin-containing monooxygenase 5 Overby, L.H. et al. (1995) Arch. Biochem. Biophys. 317:275-284.
3	7476203CD1	g2653663	5.30E-91	[Mus musculus] cytochrome P450 monooxygenase
4	7478583CD1	g14475602	3.00E-72	[fl][Bacillus cereus] acetylacetoin reductase
4	7478583CD1	g529564	2.00E-61	"[Pseudomonas putida] 2,3-butanediol dehydrogenase. Huang, M. et al. (1994) FEMS Microbiol. Lett. 124:141-150.
5	7478585CD1	g14582559	1.00E-125	[fl][Staphylococcus aureus] (AF282920) quinone oxidoreductase
6	7479904CD1	g6683558	2.30E-255	[Mus musculus] heparan sulfate 6- sulfotransferase 2
7	7480367CD1	g242749	1.30E-101	[Homo sapiens] glutathione S- transferase Hs1 subunit (EC 2.5.1.18) Rozen, F. et al. (1992) Arch. Biochem. Biophys. 292:589-593.
8	8069390CD1	g5531815	1.90E-56	[Homo sapiens] steroid dehydrogenase homolog
8	8069390CD1	g13183088	1.00E-179	[fl][Homo sapiens] steroid dehydrogenase-like protein

Table 2 (cont.)

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
9	7473869CD1	g14331125	2.00E-78	[fl][Canis familiaris] carboxylesterase D1
10	7478588CD1	g14331125	1.00E-140	[fl][Canis familiaris] carboxylesterase D1
11	55046125CD1	g6683558	1.80E-236	[Mus musculus] heparan sulfate 6- sulfotransferase 2
12	3538709CD1	g2633182	1.30E-11	[Bacillus subtilis] similar to epoxide hydrolase
12	3538709CD1	g2649734	2.00E-14	[Archaeoglobus fulgidus] carboxylesterase (est-1)
13	71563101CD1	g1944136	1.10E-234	[Rattus norvegicus] beta-alanine- pyruvate aminotransferase
14	7472027CD1	g1151081	5.40E-34	[Cavia porcellus] hydroxysteroid sulfotransferase 2
15	7480358CD1	g1066120	2.10E-25	[Cavia porcellus] copper-zinc superoxide dismutase
16	1618256CD1	g10880453	2.00E-68	[Homo sapiens] glycoprotein beta-Gal 3'-sulfotransferase
17	3387823CD1	g6782275	3.00E-99	[Caenorhabditis elegans] similar to monoamine oxidase-cDNA EST The C. elegans Sequencing Consortium. Genome sequence of the nematode C. elegans: a platform for investigating biology. Science 282:2012-2018.

Table 2 (cont.)

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
18	55142051CD1	g179955	9.50E-41	[Homo sapiens] catechol-O-methyltransferase Bertocci, B. et al. (1991) Human catechol-O-methyltransferase: cloning and expression of the membrane-associated form. Proc. Natl. Acad. Sci. U.S.A. 88:1416-1420.
19	7395274CD1	g5852342	4.60E-99	[Fungulus heteroclitus] cytochrome P450 2N1 Oleksiak, M.F. et al. (2000) Identification, functional characterization, and regulation of a new cytochrome P450 subfamily, the CYP2Ns. J. Biol. Chem. 275:2312-2321.

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	7248285CD1	756	T29, S39, T56, S85, S107, S123, S167, T239, S304, T334, S356, T386, T484, S498, S527, S531, T600, T621, T673	N198, N629	OXIDASE; LYSINE; LYSYL; COPPER DOMAIN; DM04978 Q05063 1-419: Q505-C731 LYSYL OXIDASE PROTEINLYSINE PRECURSOR SIGNAL OXIDASE OXIDOREDUCTASE COPPER GLYCOPROTEIN HOMOLOG: PD012364: P533-C731 Lysyl oxidase signature: PR00074: R593-T621, T621-C648, D679-E707, S708-N735 Signal peptide (signal peptide): M1-P24 Lysyl oxidase Lysyl oxidase: P533-S736 Signal cleavage: M1-P24	BLAST-DOMO BLAST-PRODOM BLIMPS-PRINTS HMMER HMMER-PFAM SPSCAN
2	7472835CD1	544	T30, S49, T63, S113, S179, S184, S196, S215, T216, S220, Y231, T285, S296, S304, T309, S318, T319, S338, T347, Y421, T424, T491, T505	N47, N62, N489	DIMETHYLANILINE MONOOXYGENASE (N-OXIDE-FORMING) DM02407 P49109 1-335: K3-F337 FLAVIN-CONTAINING MONOOX PR00370: A5-K20, E28-K46, F141-P157, K186-L204, N323-D341, G377-G396, Y474-R495 Pyridine nucleotide disulphide reductase class-II signature: PR00469: R187-K209, D325-Y333, F183-V207, V326-T347 Flavin-binding monooxygenase-like FMO-like: K3-F535 Pyridine nucleotide-disulphide oxidoreductases class-II active site (pyridine_redox_2.prf): P157-L214 Signal cleavage: M1-A18	BLAST-DOMO BLIMPS-PRINTS BLIMPS-PRINTS HMMER-PFAM PROFILES SCAN SPSCAN
3	7476203CD1	501	S54, T118, T174, S372, T402, S412, T454		CYTOCHROME P450 DM00022 P52786 83-492: S71-P478 CYTOCHROME P450 MONOOXYGENASE OXIDOREDUCTASE HEME ELECTRON TRANSPORT MEMBRANE MICROSOME ENDOPLASMIC: PD000021: Q237-L351, D396-P464, P31-K60, E339-F411, N108-E158	BLAST-DOMO BLAST-PRODOM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
3	(cont)				P450 superfamily signature: PR00385: A349-V360, L425-Y434, G296-I313, Q314-D327 Mitochondrial P450 signature: PR00408: S114-R124, G296-I313, K342-V360 E-class P450 group I: PR00463: S173-D191, N285-A302, I305-G331, M348-V366, N389-D413, F424-Y434, Y434-F457 Signal peptide (signal_peptide): M1-C24 Transmembrane domain (transmem_domain): L4-F23, L197-Q214 Cytochrome P450 (p450): P31-M67, R107-A485 Cytochrome P450 cysteine heme-iron ligand signature (cytochrome_p450.prf): F406-R456 Signal cleavage: M1-Q49 ZINC-CONTAINING ALCOHOL DEHYDROGENASES DM00064 P39714 18-373: E105-I343, E21-S99 OXIDOREDUCTASE ZINC DEHYDROGENASE ALCOHOL NAD PROTEIN FAMILY MULTIGENE NADP FORMALDEHYDE: PD000104: A3-V265, G168-A246 Zinc-containing alcohol dehydrogenase proteins: BL00059: E26-H42, P65-V92, G117-Q158, G168-T214 Zinc-binding dehydrogenases (adh_zinc): K2-K344 Adh_zinc G69-V83 Zinc-containing alcohol dehydrogenase signatures (adh_zinc.prf): G38-E102 ZINC-CONTAINING ALCOHOL DEHYDROGENASES DM00064 S57611 3-340: Q32-V356 Zinc-binding dehydrogenases (adh_zinc): D46-K357 (E-value = 4.4e-06, Score = -68.6) Quinone oxidoreductase: BL01162: S150-K193 (p < 0.057)	BLIMPS-PRINTS BLIMPS-PRINTS BLIMPS-PRINTS HMMER HMMER HMMER-PFAM PROFILES SCAN SPSCAN BLAST-DOMO BLAST-PRODOM BLIMPS-BLOCKS HMMER-PFAM MOTIFS PROFILES SCAN BLAST-DOMO HMMER-PFAM BLIMPS-BLOCKS
4	7478583CD1	345	S143 S219 S251 S327 S99 T248 T53	N141		
5	7478585CD1	361	T14, Y86, S107, N81 S113, T330			

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
6	7479904CD1	499	S5, S50, S55, T65, T120, T129, S156, S187, S199, T205, S208, T237, T251, T260, Y277, S314, T355, S359, S464, S471, S476	N63, N204, N298, N354, N438, N450, N458, N483, N486	HEPARANSULFATE 6-SULFOTRANSFERASE TRANSFERASE: PD139786: K207-R412, M1-K182 Signal peptide (signal_peptide): M1-Q24 Transmembrane domain (transmem_domain): L8-Y25 Signal cleavage: M1-Q24	BLAST-PRODOM HMMER HMMER SPSCAN
7	7480367CD1	222			GLUTATHIONE TRANSFERASE DM00127 P08263 71-190: L72-P192 GLUTATHIONE TRANSFERASE STRANSFERASE MULTIGENE FAMILY PROTEIN CLASS-ALPHA S-CRYSTALLIN GST LYASE: PD000312: L81-R204, K6-L102 (p = 8.0e-10) Glutathione S-transferase: PF00043: Q54-G83 Glutathione S-transferases: L7-P192 SHORT-CHAIN ALCOHOL DEHYDROGENASE FAMILY DM00034 F37058 43-275: I63-I297 Short-chain dehydrogenas: BL00061: G198-E235 Alcohol dehydrogenase: PR00080: F218-A237, K143-V154, G198-G206 Glucose/ribitol dehydrogenase: PR00081: W69-E86, K143-V154, M192-C208, F218-A237, K239-T256 Short chain dehydrogenase (adh_short): R68-S254 Short-chain alcohol dehydrogenase family signature (adh_short.prif): G198-P248 Signal cleavage: M1-K36	BLAST-DOMO BLAST-PRODOM BLIMPS-PRINTS HMMER-PFAM BLAST-DOMO BLIMPS-BLOCKS BLIMPS-PRINTS BLIMPS-PRINTS HMMER-PFAM PROFILES-SCAN SPSCAN
8	8069390CD1	330	T33, S58, S98, T114, S128, S271, S286, T290, S320	N318		

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
9	7473869CD1	303	S4, Y20, S52, S81, S199, T277, S284	N116, N178	CARBOXYLESTERASES TYPE-B: DM00175 P16303 20-361: ML-M264 ESTERASE HYDROLASE PRECURSOR SIGNAL GLYCOPROTEIN SERINE PROTEIN CARBOXYLESTERASE FAMILY MULTIGENE PD000169: S4-M281 Carboxylesterases type-B: BL00122: E16-P26, P36-F46, V65-L80, G89-I129, L141-G150 Carboxylesterases (Coesterase): L12-F266 Carboxylesterase motifs: (carboxylesterase_B.1): F110-G125 (carboxylesterase_B.2): E16-P26 Carboxylesterases type-B signatures (carboxylesterase_b.pr.f): N90-A145 CARBOXYLESTERASES TYPE-B DM00175 Q04791 26-353: K48-M385 ESTERASE HYDROLASE PRECURSOR SIGNAL GLYCOPROTEIN SERINE PROTEIN CARBOXYLESTERASE FAMILY MULTIGENE: PD000169: M37-E430, W26-T203, M394-W569, A488-P578 Carboxylesterases type-B: BL00122: F74-P94, E137-P147, P157-F167, V186-L201, G210-I250, L262-G271, W522-N532 Carboxylesterases Coesterase: R25-W569 Carboxylesterase_B.1: F231-G246; Carboxylesterase_B.2: E137-P147 Carboxylesterases type-B signatures (carboxylesterase_b.pr.f): N211-A266 Signal cleavage: M1-G43	BLAST-DOMO BLAST-PRODOM BLIMPS-BLOCKS HMER-PFAM MOTIFS PROFILES SCAN BLAST-DOMO BLAST-PRODOM BLIMPS-BLOCKS HMER-PFAM MOTIFS PROFILES SCAN SPSCAN
10	7478588CD1	584	S23, T47, Y107, S125, Y141, S173, S202, S320, T398, S405, Y548, T554, S575	N237, N299, N411		

Table 3 (cont.)

SEQ ID NO.	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
11	55046125CD1	508	S28, S54, S99, S104, T114, T169, T178, S205, T246, T260, T269, Y286, S323, T364, S368, S473, S480, S485,	N112, N307, N363, N447, N459, N467, N492, N495	HEPARANSULFATE 6-SULFOTRANSFERASE TRANSFERASE: PD139786:M50-R421 Transmembrane domain (transmem_domain): L57-Y74	BLAST-PRODOM HMME
12	3538709CD1	439	T34, T83, T167, S168, T281, S423	N108, N332	Alpha/beta hydrolase fold: PR00111: G202-S217, A262-I275, V362-E376 Epoxide hydrolase signature: PR00412: H182-R200, G202-S217, I248-L261, I393-F415 Alpha/beta hydrolase fold (abhydrolase): Y203-L416	BLIMPS-PRINTS BLIMPS-PRINTS HMME-PFAM
13	71563101CD1	514	T32, S169, S192, T248, S253, T363, S448, T480, T490		AMINOTRANSFERASES CLASS-III PYRIDOXAL-PHOSPHATE ATTACHMENT DM00188 P16932 2-426: D271-A504, K84-R233 ALANINEGLYOXYLATE AMINOTRANSFERASE 2 PRECURSOR EC 2.6.1.44 AGT BETAALANINEPYRUVATE BETAALAT II TRANSFERASE PYRIDOXAL PHOSPHATE MITOCHONDRION TRANSIT PEPTIDE: PD032383: M1-E94 Aminotransferases class-3: BL00600: P343-G355, H377-L395, D98-P121, P160-S185, I190-L205, I281-G294, F302-G330 Transmembrane domain: Y104-N124 Aminotransferases class-III pyridoxal-phosphate: aminotran_3: W95-T208 I281-M487 Aminotransferases class-III pyridoxal-phosphate attachment site (aa_transfer_class_3.prp): E308-C374	BLAST-DOMO BLAST-PRODOM BLIMPS-BLOCKS HMME HMME-PFAM PROFILES-SCAN

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
14	7472027CD1	226	S17, T23, T28, S106, S136, T195	N21, N66	PHOSPHADENOSINE-PHOSPHOSULFATE SYNTHETASE (PAPS) BINDING SITE BINDING DM00981 JC4531 1-287: M50-E193 TRANSFERASE SULFOTRANSFERASE STEROID METABOLISM HYDROXYSTEROID ALCOHOL PHENOL ESTROGEN PROTEIN STEROIDBINDING PD001218: L73-Y165	BLAST-DOMO  BLAST-PRODOM
15	7480358CD1	121	T38, S96, T119	N69	COPPER/ZINC SUPEROXIDE DISMUTASE DM00227 P00442 3-147: K3-V116 SUPEROXIDE DISMUTASE CUZN OXIDOREDUCTASE COPPER ZINC PRECURSOR SIGNAL PERIPLASMIC EXTRACELLULAR: PD000469: M2-V116 Copper/zinc superoxide dismutase proteins: BL00087: K3-Q22, R40-W83 Cu-Zn-superoxide dismutase: PR00068: R43-P65, V97-T119 Copper/zinc superoxide dismutase (SODC): (sodcu): M2-P73, S96-V116 Signal peptide: M1-G29	BLAST-DOMO  BLAST-PRODOM  BLIMPS-BLOCKS BLIMPS-PRINTS
16	1618256CD1	486	T9, S55, T74, T265, S296, T349, T356, T420, S431, S444, T480	N374	CEREBROSIDE SULFOTRANSFERASE: PD140372: W13-G215, P279-Y461	HMMER-PFAM SIGPEPT SPSCAN BLAST-PRODOM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
17	3387823CD1	649	T64, T184, T236, S253, S368, S382, S431, T438, S471, S534, S356, S516, S547, T550, T563, T624	N209	PROTEIN SIMILARITY KINASE SLIME MOLD R13G10.2 CHROMOSOME III OXIDOREDUCTASE FLAVOPROTEIN: PD041935: L358-D433, D250-G337 Signal peptide: M1-E22 Flavin containing amine oxidase: Amino oxidase: P219-A647 Adrenodoxin reductase family (flavin-containing component of mitochondrial monooxygenase system): PR00419A: S211-I233, PR00419B: K234-R247, PR00419D: V212-R226 Aromatic-ring hydroxylase (flavoprotein monooxygenase): PR00420A: S211-I233, PR00420C: Q614-P629	BLAST-PRODOM  SIGPEPT HMMEP-PFAM BLIMPS-PRINTS BLIMPS-PRINTS
18	55142051CD1	258	S35, T67, S73, S74, S159, S174		Transmembrane domain: M1-V18 O-METHYLTRANSFERASE TRANSFERASE METHYLTRANSFERASE: PD173699: I65-D162 (P=9.5e-10) CATECHOL-O-METHYLTRANSFERASE CHAIN: DM02481 P22734 I41-139: E46-A137	HMMEP BLAST-PRODOM BLAST-DOMO

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
19	7395274CD1	544	S85, S88, S135, S148, S181, T184, Y205, T244, S335, S336, S393, T395, T406, T479, S519, T525, S542	S331, N334, N538	CYTOCHROME P450 DM00022 P10635 52-480: L110-F528 CYTOCHROME P450 MONOOXYGENASE OXIDOREDUCTASE HEME ELECTRON TRANSPORT MEMBRANE MICROSOME ENDOPLASMIC PD000021: K285-G433, I386-Q471, H450-I539, I436-I539, A227-D338, G58-I231 Cytochrome P450 cysteine BL00086: F480-F511 Mitochondrial P450 signa PR00408:I370-E383, A398-V416, F462-D470, I481-C490, C490-F501, W50-P65, H169-K179, A352-S369 E-class P450 group I PR00463: A111-L130, S135-I156, A227-N245, Y341-T358, L361-G387, E404-P422, N445-D469, F480-C490, C490-F513 E-class P450 group IV PR00465: V105-V128, H450-L468, L474-C490, C490-M508 Cytochrome P450: L107-R543 Cytochrome P450: F483-G492 Cytochrome P450 cysteine heme-iron ligand signature (cytochrome_p450.prf): P464-A512 Signal cleavage: M1-G47	BLAST-DOMO BLAST-PRODOM BLIMPS-PRINTS BLIMPS-PRINTS BLIMPS-PRINTS HMMER-PFAM MOTIFS PROFILES SCAN SPSCAN

Table 4

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
20	7248285CB1	2603	2496-2603, 526-572	7690384J1 (PROSTME06) 6460335P7 (OSTEUNC01) 8035282H1 (SMCRUNE01) 2796172F6 (NPOLNOT01) 7690384H1 (PROSTME06) FL7248285_g7939436_000006 _g1890108_1_7-8 GNN:g7939436_000006_002.e dit	652 1144 276 2072 1650 1106 1 1	1330 1999 892 2603 2185 1428 456
21	7472833CB1	1745	1-57, 412- 1745	GBI.g7107731_0001.edit_2. comp	1445	1745
22	7476203CB1	1587	1-209, 280- 1587	GBI.g7107731_0001.edit_1 GBI.g8567322_000003_00000 4.edit GBI.g8567322_000003_00000 4.edit_2	1 1 960	1674 1139 1587
23	7478583CB1	1038	1-1038	GNN.g8576127_000058_002	1	1038
24	7478585CB1	1584	1-1584	GNN.g8576127_CDS_1.edit GNN.7478585CT1	1 596	1467 1584
25	7479904CB1	1500	1-149	g1655588_g3191972_g856919 3.edit	1	1500
26	7480367CB1	669		GBI:g8348076_edit_1 2090065H1 (PANCNOT04)	1 508	669 669

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
27	8069390CB1	3551	1609-1783, 701-898, 1- 21, 3101-3551	7360250H1 (BRAIFEE05) 7660957H1 (OVARNOE02) 7660957J1 (OVARNOE02) 6450195H1 (BRAINOC01) 1759470R6 (PITUNOT03) 1619329T6 (BRAITUT12) 4249653R6 (BRADDI01) 91203094 1576067T6 (LNODNOT03) 7365248H1 (OVARDIC01) 4249653F6 (BRADDI01) 7100733H1 (BRAWTDR02) 3333887F6 (BRAIFET01) 55023671J1 55051441J1 55071187J1 55071183H1 55017748J1 7580667H1 (BRAIFEC01) 60251275D1 55058627J1 6205079H1 (PITUNON01) 4365390H1 (SKIRNOT01) 55111596H1 55071187J1 71975318V1 2783360F6 (BRSTNOT13) 55071237J1 55046165J2 55046125H1 5626585H1 (PLACFER01)	607 2157 1 330 2966 2859 2357 3095 2736 920 1748 1597 1068 1360 1 1045 987 1496 543 1333 157 1562 1129 1 752 1694 2121 1 720 314 1512	1056 2645 537 989 3512 3484 2781 3551 3390 1560 2196 2148 1673 2178 1032 1662 1606 1970 1108 1837 714 2081 1391 540 1365 2315 2642 548 1572 1143 2108
28	7473869CB1	2178	1-231, 1549- 1579, 1949- 2178, 654-732			
29	7478588CB1	2081	1-97, 1256- 1450			
30	55046125CB1	2642	2250-2642, 403-468			

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
31	3538709CB1	2080	863-1020, 2027-2080, 580-639	7067524H1 (BRATNOR01) 2225614F6 (SEAVNOT01) 1668310H1 (EMARNOT03) 3763474H1 (BRAINOT21) 7764242H1 (URETTUE01) 6883439J1 (BRAHTDR03) 7191721H2 (BRATDIO01) 2560864H1 (ADRETUT01)	624 1797 1141 1166 1 1315 253 1656	1178 2080 1376 1461 463 1932 855 1947
32	71563101CB1	2219	1-44, 541- 1063	92985189 71563534V1 71562178V1 6926859H1 (SINLTM01) 293718H1 (LIVRN04) 71563101V1 FL7472027CT1_00001 GNN.g9211248_000026_002	1734 684 1184 1 1828 529 1 1	2212 1366 1840 662 2066 1354 681 399
33	7472027CB1	681	148-681		1	681
34	7480358CB1	399	1-35		1	399
35	1618256CB1	2302	2241-2302, 1- 260, 1172- 1194	7010610H1 (COLNEEC01) 1618256F6 (BRATUT12) 6330412F8 (BRANDIN01) 7403247H1 (SINIDME01) 4922615F9 (TESTNOT11) 2605131F6 (LUNGUT07)	1 1329 359 1475 646 1835	604 1876 1086 2112 1374 2302

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
36	3387823CB1	3341	1-447, 3270- 3341, 878- 1415	7741282H1 (THYMNOE01)	1	662
				72335796V1	2819	3341
				70858392V1	1919	2549
				70857153V1	2459	2872
				70856101V1	1381	1902
				657765H1 (EOSINOT03)	2322	2590
				70002313D1	1551	2004
				995938H1 (KIDNTUT01)	2584	2902
				677154J1 (BRAUNOR01)	768	1218
				70857616V1	1116	1524
				7087862H1 (BRAUTDR03)	438	906
				55142051J1	72	337
37	55142051CB1	777	1-257	GBI-g8567879 edit	1	777
38	7395274CB1	3600	1-1390, 2414- 3076, 2342- 2372, 3576- 3600	6779264H1 (OVARDIR01)	2278	2967
				g2064300	3001	3596
				55052831H1	787	1597
				7394114R8 (BRABDIE02)	454	1371
				8128245H1 (SCOMDIE01)	1	693
				2300407R6 (BRSTNOT05)	2076	2545
				2300407T6 (BRSTNOT05)	3044	3568
				7996595H1 (ADRETUC01)	1493	2221
				g2079829	3243	3600
				7458912H1 (LIVRTUE01)	1677	2255
				1259839T6 (MENITUT03)	2869	3567
				6544865H1 (LNODNON02)	3124	3590

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID	Representative Library
20	7248285CB1	PROSTME06
26	7480367CB1	SINTNOT02
27	8089390CB1	HNT2AGT01
28	7473869CB1	BRAQTR02
29	7478588CB1	OVARDIR01
30	55046125CB1	BRSTNOT13
31	3538709CB1	BRAHTR03
32	71563101CB1	SINTMR01
35	1618256CB1	BRAITUT12
36	3387823CB1	BRAINOT19
38	7395274CB1	BRABDIE02

Table 6

Library	Vector	Library Description
BRABDIE02	pINCY	This 5' biased random primed library was constructed using RNA isolated from diseased cerebellum tissue removed from the brain of a 57-year-old Caucasian male who died from a cerebrovascular accident. Serologies were negative. Patient history included Huntington's disease, emphysema, and tobacco abuse (3-4 packs per day, for 40 years).
BRAHTR03	PCDNA2.1	This random primed library was constructed using RNA isolated from archaer cortex, anterior hippocampus tissue removed from a 55-year-old Caucasian female who died from cholangiocarcinoma. Pathology indicated mild meningeal fibrosis predominately over the convexities, scattered axonal spheroids in the white matter of the cingulate cortex and the thalamus, and a few scattered neurofibrillary tangles in the entorhinal cortex and the periaqueductal gray region. Pathology for the associated tumor tissue indicated well-differentiated cholangiocarcinoma of the liver with residual or relapsed tumor. Patient history included cholangiocarcinoma, post-operative Budd-Chiari syndrome, biliary ascites, hydorthorax, dehydration, malnutrition, oliguria and acute renal failure. Previous surgeries included cholecystectomy and resection of 85% of the liver.
BRAINOT19	pINCY	Library was constructed using RNA isolated from diseased brain tissue removed from the left frontal lobe of a 27-year-old Caucasian male during a brain lobectomy. Pathology indicated a focal deep white matter lesion, characterized by marked gliosis, calcifications, and hemosiderin-laden macrophages, consistent with a remote perinatal injury. This tissue also showed mild to moderate generalized gliosis, predominantly subpial and subcortical, consistent with chronic seizure disorder. The left temporal lobe, including the mesial temporal structures, showed focal, marked pyramidal cell loss and gliosis in hippocampal sector CA1, consistent with mesial temporal sclerosis. GFAP was positive for astrocytes. The patient presented with intractable epilepsy, focal epilepsy, hemiplegia, and an unspecified brain injury. Patient history included cerebral palsy, abnormality of gait, and depressive disorder. Family history included brain cancer.

Table 6 (cont.)

Library	Vector	Library Description
BRAITUT12	PINCY	Library was constructed using RNA isolated from brain tumor tissue removed from the left frontal lobe of a 40-year-old Caucasian female during excision of a cerebral meningioma lesion. Pathology indicated grade 4 gemistocytic astrocytoma.
BRAQTR02	PCDNA2.1	This random primed library was constructed using RNA isolated from mixed midbrain tissues (nucleus raphe magnus, periaqueductal gray and substantia nigra) removed from a 55-year-old Caucasian female who died from cholangiocarcinoma. Pathology indicated mild meningioma fibrosis predominately over the convexities, scattered axonal spheroids in the white matter of the cingulate cortex and the thalamus, and a few scattered neurofibrillary tangles in the entorhinal cortex and the periaqueductal gray region. Pathology for the associated tumor tissue indicated well-differentiated cholangiocarcinoma of the liver with residual or relapsed tumor. Patient history included cholangiocarcinoma, post-operative Budd-Chiari syndrome, biliary ascites, hydrothorax, dehydration, malnutrition, oliguria and acute renal failure. Previous surgeries included cholecystectomy and resection of 85% of the liver.
BRSTNOT13	PINCY	Library was constructed using RNA isolated from breast tissue removed from a 36-year-old Caucasian female during bilateral simple mastectomy. Patient history included a breast neoplasm, depressive disorder, hyperlipidemia, and a chronic stomach ulcer. Family history included a cardiovascular and cerebrovascular disease, hyperlipidemia, skin, breast, esophageal, bladder, and bone cancer, and Hodgkin's lymphoma.
HNT2AGT01	PBLUESCRIPT	Library was constructed at Stratagene (STR937233), using RNA isolated from the hNT2 cell line derived from a human teratocarcinoma that exhibited properties characteristic of a committed neuronal precursor. Cells were treated with retinoic acid for 5 weeks and with mitotic inhibitors for two weeks and allowed to mature for an additional 4 weeks in conditioned medium.
OVARDIR01	PCDNA2.1	This random primed library was constructed using RNA isolated from right ovary tissue removed from a 45-year-old Caucasian female during total abdominal hysterectomy, bilateral salpingo-oophorectomy, vaginal suspension and fixation, and incidental appendectomy. Pathology indicated stromal hyperthecosis of the right and left ovaries. Pathology for the matched tumor tissue indicated a dermoid cyst (benign cystic teratoma) in the left ovary. Multiple (3) intramural leiomyomata were identified. The cervix showed squamous metaplasia. Patient history included metrorrhagia, female stress incontinence, alopecia, depressive disorder, pneumonia, normal delivery, and deficiency anemia. Family history included benign hypertension, atherosclerotic coronary artery disease, hyperlipidemia, and primary tuberculous complex.

Table 6 (cont.)

Library	Vector	Library Description
PROSTM06	PCDNA2.1	<p>This 5' biased random primed library was constructed using RNA isolated from diseased prostate tissue removed from a 57-year-old Caucasian male during closed prostatic biopsy, radical prostatectomy, and regional lymph node excision. Pathology indicated adenofibromatous hyperplasia. Pathology for the matched tumor tissue indicated adenocarcinoma, Gleason grade 3+3, forming a predominant mass involving the right side centrally. The patient presented with elevated prostate specific antigen and prostate cancer. Patient history included tobacco abuse in remission. Previous surgeries included cholecystectomy, repair of diaphragm hernia, and repair of vertebral fracture. Patient medications included Pepsid, Omnipen, and Eulexin. Family history included benign hypertension, cerebrovascular accident, atherosclerotic coronary artery disease, uterine cancer and type II diabetes in the mother; prostate cancer in the father; drug abuse, prostate cancer, and breast cancer in the sibling(s).</p>
SINITMR01	PCDNA2.1	<p>This random primed library was constructed using RNA isolated from ileum tissue removed from a 70-year-old Caucasian female during right hemicolectomy, open liver biopsy, flexible sigmoidoscopy, colonoscopy, and permanent colostomy. Pathology for the matched tumor tissue indicated invasive grade 2 adenocarcinoma forming an ulcerated mass, situated 2 cm distal to the ileocecal valve. Patient history included a malignant breast neoplasm, type II diabetes, hyperlipidemia, viral hepatitis, an unspecified thyroid disorder, osteoarthritis, a malignant skin neoplasm, deficiency anemia, and normal delivery. Family history included breast cancer, atherosclerotic coronary artery disease, benign hypertension, cerebrovascular disease, ovarian cancer, and hyperlipidemia.</p>
SINITMT02	pBLUESCRIPT	<p>Library was constructed using RNA isolated from the small intestine of a 55-year-old Caucasian female, who died from a subarachnoid hemorrhage. Serologies were positive for cytomegalovirus (CMV). Previous surgeries included a hysterectomy.</p>

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL.PDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	ESTs: Probability value= 1.0E-8 or less Full Length sequences: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E value=1.0E-6 Assembled ESTs: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less Full Length sequences: fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Atwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	PFAM hits: Probability value= 1.0E-3 or less Signal peptide hits: Score= 0 or greater

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribkov, M. et al. (1988) CABIOS 4:61-66; Gribkov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score $\geq$ GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:
  - a) a polypeptide comprising an amino acid sequence selected from the group consisting of  
5 SEQ ID NO:1-19 ,
  - b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-19 ,
  - c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19 , and
  - 10 d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19 .
2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1-19 .
- 15 3. An isolated polynucleotide encoding a polypeptide of claim 1.
4. An isolated polynucleotide encoding a polypeptide of claim 2.
5. An isolated polynucleotide of claim 4 selected from the group consisting of SEQ ID  
20 NO:20-38.
6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.
- 25 7. A cell transformed with a recombinant polynucleotide of claim 6.
8. A transgenic organism comprising a recombinant polynucleotide of claim 6.
9. A method of producing a polypeptide of claim 1, the method comprising:
  - 30 a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and
  - b) recovering the polypeptide so expressed.

10. An isolated antibody which specifically binds to a polypeptide of claim 1.
11. An isolated polynucleotide selected from the group consisting of:
- 5 a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:20-38,
- b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:20-38,
- c) a polynucleotide complementary to a polynucleotide of a),
- d) a polynucleotide complementary to a polynucleotide of b), and
- 10 e) an RNA equivalent of a)-d).
12. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 11.
- 15 13. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:
- a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex
- 20 is formed between said probe and said target polynucleotide or fragments thereof, and
- b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.
14. A method of claim 13, wherein the probe comprises at least 60 contiguous nucleotides.
- 25 15. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:
- a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and
- 30 b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.
16. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable

excipient.

17. A composition of claim 16, wherein the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO:1-19 .

5

18. A method for treating a disease or condition associated with decreased expression of functional DME, comprising administering to a patient in need of such treatment the composition of claim 16.

10

19. A method of screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- b) detecting agonist activity in the sample.

15

20. A composition comprising an agonist compound identified by a method of claim 19 and a pharmaceutically acceptable excipient.

20

21. A method for treating a disease or condition associated with decreased expression of functional DME, comprising administering to a patient in need of such treatment a composition of claim 20.

22. A method of screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

25

- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- b) detecting antagonist activity in the sample.

23. A composition comprising an antagonist compound identified by a method of claim 22 and a pharmaceutically acceptable excipient.

30

24. A method for treating a disease or condition associated with overexpression of functional DME, comprising administering to a patient in need of such treatment a composition of claim 23.

25. A method of screening for a compound that specifically binds to the polypeptide of claim

1, the method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and

b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a  
5 compound that specifically binds to the polypeptide of claim 1.

26. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, the method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under conditions  
10 permissive for the activity of the polypeptide of claim 1,

b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound,  
and

c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound  
with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change  
15 in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a  
compound that modulates the activity of the polypeptide of claim 1.

27. A method of screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method  
20 comprising:

a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,

b) detecting altered expression of the target polynucleotide, and

c) comparing the expression of the target polynucleotide in the presence of varying amounts  
25 of the compound and in the absence of the compound.

28. A method of assessing toxicity of a test compound, the method comprising:

a) treating a biological sample containing nucleic acids with the test compound,

b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at  
30 least 20 contiguous nucleotides of a polynucleotide of claim 11 under conditions whereby a specific  
hybridization complex is formed between said probe and a target polynucleotide in the biological  
sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim  
11 or fragment thereof,

- c) quantifying the amount of hybridization complex, and
- d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

5

29. A diagnostic test for a condition or disease associated with the expression of DME in a biological sample, the method comprising:

- a) combining the biological sample with an antibody of claim 10, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex, and
- b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

10

30. The antibody of claim 10, wherein the antibody is:

- a) a chimeric antibody,
- b) a single chain antibody,
- c) a Fab fragment,
- d) a F(ab')<sub>2</sub> fragment, or
- e) a humanized antibody.

15

31. A composition comprising an antibody of claim 10 and an acceptable excipient.

20

32. A method of diagnosing a condition or disease associated with the expression of DME in a subject, comprising administering to said subject an effective amount of the composition of claim 31.

33. A composition of claim 31, wherein the antibody is labeled.

25

34. A method of diagnosing a condition or disease associated with the expression of DME in a subject, comprising administering to said subject an effective amount of the composition of claim 33.

35. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 10, the method comprising:

30

- a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, or an immunogenic fragment thereof, under conditions to elicit

an antibody response,

b) isolating antibodies from said animal, and

c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19 .

36. An antibody produced by a method of claim 35.

37. A composition comprising the antibody of claim 36 and a suitable carrier.

38. A method of making a monoclonal antibody with the specificity of the antibody of claim 10, the method comprising:

a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19 , or an immunogenic fragment thereof, under conditions to elicit an antibody response,

b) isolating antibody producing cells from the animal,

c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells,

d) culturing the hybridoma cells, and

e) isolating from the culture monoclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19 .

39. A monoclonal antibody produced by a method of claim 38.

40. A composition comprising the antibody of claim 39 and a suitable carrier.

41. The antibody of claim 10, wherein the antibody is produced by screening a Fab expression library.

42. The antibody of claim 10, wherein the antibody is produced by screening a recombinant immunoglobulin library.

43. A method of detecting a polypeptide having an amino acid sequence selected from the

group consisting of SEQ ID NO:1-19 in a sample, the method comprising:

a) incubating the antibody of claim 10 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and

b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide  
5 having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19 in the sample.

44. A method of purifying a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19 from a sample, the method comprising:

a) incubating the antibody of claim 10 with a sample under conditions to allow specific binding  
10 of the antibody and the polypeptide, and

b) separating the antibody from the sample and obtaining the purified polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19 .

45. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

15

46. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

47. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

20

48. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

49. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

50. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

25

51. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

52. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

30

53. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

54. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.

55. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.

56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.

5 57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:13.

58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:14.

59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:15.

10

60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:16.

61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:17.

15

62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:18.

63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:19.

64. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:20.

20

65. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:21.

66. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:22.

25

67. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:23.

68. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:24.

30

69. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:25.

70. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:26.

71. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:27.

72. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:28.

73. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:29.

5 74. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:30.

75. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:31.

10 76. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:32.

77. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:33.

78. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:34.

15 79. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:35.

80. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:36.

20 81. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:37.

82. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:38.

## &lt;110&gt; INCYTE GENOMICS, INC.

BAUGHN, Mariah R.  
 BRUNS, Christopher M.  
 DAS, Debopriya  
 DELEGEANE, Angelo M.  
 DING, Li  
 ELLIOT, Vicki S.  
 GANDHI, Ameena R.  
 GRIFFIN, Jennifer A.  
 HAFALIA, April J.A.  
 KHAN, Farrah A.  
 LAL, Preeti  
 LEE, Sally  
 LU, Dyung Aina M.  
 LU, Yan  
 PATTERSON, Chandra  
 RAMKUMAR, Jayalaxmi  
 RING, Huijun Z.  
 SANJANWALA, Madhu S.  
 TANG, Y. Tom  
 THANGAVELU, Kavitha  
 THORNTON, Michael  
 TRIBOULEY, Catherine M.  
 WALIA, Narinder K.  
 WARREN, Bridget A.  
 YANG, Junming  
 YAO, Monique G.  
 YUE, Henry

## &lt;120&gt; DRUG METABOLIZING ENZYMES

## &lt;130&gt; PI-0185 PCT

## &lt;140&gt; To Be Assigned

## &lt;141&gt; Herewith

<150> 60/223,055; 60/224,728; 60/226,440; 60/228,067; 60/230,063; 60/232,244;  
 60/234,269

<151> 2000-08-04; 2000-08-11; 2000-08-18; 2000-08-24; 2000-08-31; 2000-09-13;  
 2000-09-20

## &lt;160&gt; 38

## &lt;170&gt; PERL Program

## &lt;210&gt; 1

## &lt;211&gt; 756

## &lt;212&gt; PRT

## &lt;213&gt; Homo sapiens

## &lt;220&gt;

## &lt;221&gt; misc\_feature

## &lt;223&gt; Incyte ID No: 7248285CD1

## &lt;400&gt; 1

Met	Ala	Trp	Ser	Pro	Pro	Ala	Thr	Leu	Phe	Leu	Phe	Leu	Leu	Leu
1				5				10					15	
Leu	Gly	Gln	Pro	Pro	Pro	Ser	Arg	Pro	Gln	Ser	Leu	Gly	Thr	Thr
				20				25					30	
Lys	Leu	Arg	Leu	Val	Gly	Pro	Glu	Ser	Lys	Pro	Glu	Glu	Gly	Arg
				35				40					45	
Leu	Glu	Val	Leu	His	Gln	Gly	Gln	Trp	Gly	Thr	Val	Cys	Asp	Asp
				50				55					60	

Asn	Phe	Ala	Ile	Gln	Glu	Ala	Thr	Val	Ala	Cys	Arg	Gln	Leu	Gly	
				65						70					75
Phe	Glu	Ala	Ala	Leu	Thr	Trp	Ala	His	Ser	Ala	Lys	Tyr	Gly	Gln	
				80						85					90
Gly	Glu	Gly	Pro	Ile	Trp	Leu	Asp	Asn	Val	Arg	Cys	Val	Gly	Thr	
				95						100					105
Glu	Ser	Ser	Leu	Asp	Gln	Cys	Gly	Ser	Asn	Gly	Trp	Gly	Val	Ser	
				110						115					120
Asp	Cys	Ser	His	Ser	Glu	Asp	Val	Gly	Val	Ile	Cys	His	Pro	Arg	
				125						130					135
Arg	His	Arg	Gly	Tyr	Leu	Ser	Glu	Thr	Val	Ser	Asn	Ala	Leu	Gly	
				140						145					150
Pro	Gln	Gly	Gln	Arg	Leu	Glu	Glu	Val	Arg	Leu	Lys	Pro	Ile	Leu	
				155						160					165
Ala	Ser	Ala	Lys	Gln	His	Ser	Pro	Val	Thr	Glu	Gly	Ala	Val	Glu	
				170						175					180
Val	Lys	Tyr	Glu	Gly	His	Trp	Arg	Gln	Val	Cys	Asp	Gln	Gly	Trp	
				185						190					195
Thr	Met	Asn	Asn	Ser	Arg	Val	Val	Cys	Gly	Met	Leu	Gly	Phe	Pro	
				200						205					210
Ser	Glu	Val	Pro	Val	Asp	Ser	His	Tyr	Tyr	Arg	Lys	Val	Trp	Asp	
				215						220					225
Leu	Lys	Met	Arg	Asp	Pro	Lys	Ser	Arg	Leu	Lys	Ser	Leu	Thr	Asn	
				230						235					240
Lys	Asn	Ser	Phe	Trp	Ile	His	Gln	Val	Thr	Cys	Leu	Gly	Thr	Glu	
				245						250					255
Pro	His	Met	Ala	Asn	Cys	Gln	Val	Gln	Val	Ala	Pro	Ala	Arg	Gly	
				260						265					270
Lys	Leu	Arg	Pro	Ala	Cys	Pro	Gly	Gly	Met	His	Ala	Val	Val	Ser	
				275						280					285
Cys	Val	Ala	Gly	Pro	His	Phe	Arg	Pro	Pro	Lys	Thr	Lys	Pro	Gln	
				290						295					300
Arg	Lys	Gly	Ser	Trp	Ala	Glu	Glu	Pro	Arg	Val	Arg	Leu	Arg	Ser	
				305						310					315
Gly	Ala	Gln	Val	Gly	Glu	Gly	Arg	Val	Glu	Val	Leu	Met	Asn	Arg	
				320						325					330
Gln	Trp	Gly	Thr	Val	Cys	Asp	His	Arg	Trp	Asn	Leu	Ile	Ser	Ala	
				335						340					345
Ser	Val	Val	Cys	Arg	Gln	Leu	Gly	Phe	Gly	Ser	Ala	Arg	Glu	Ala	
				350						355					360
Leu	Phe	Gly	Ala	Arg	Leu	Gly	Gln	Gly	Leu	Gly	Pro	Ile	His	Leu	
				365						370					375
Ser	Glu	Val	Arg	Cys	Arg	Gly	Tyr	Glu	Arg	Thr	Leu	Ser	Asp	Cys	
				380						385					390
Pro	Ala	Leu	Glu	Gly	Ser	Gln	Asn	Gly	Cys	Gln	His	Glu	Asn	Asp	
				395						400					405
Ala	Ala	Val	Arg	Cys	Asn	Val	Pro	Asn	Met	Gly	Phe	Gln	Asn	Gln	
				410						415					420
Val	Arg	Leu	Ala	Gly	Gly	Arg	Ile	Pro	Glu	Glu	Gly	Leu	Leu	Glu	
				425						430					435
Val	Gln	Val	Glu	Val	Asn	Gly	Val	Pro	Arg	Trp	Gly	Ser	Val	Cys	
				440						445					450
Ser	Glu	Asn	Trp	Gly	Leu	Thr	Glu	Ala	Met	Val	Ala	Cys	Arg	Gln	
				455						460					465
Leu	Gly	Leu	Gly	Phe	Ala	Ile	His	Ala	Tyr	Lys	Glu	Thr	Trp	Phe	
				470						475					480
Trp	Ser	Gly	Thr	Pro	Arg	Ala	Gln	Glu	Val	Val	Met	Ser	Gly	Val	
				485						490					495
Arg	Cys	Ser	Gly	Thr	Glu	Leu	Ala	Leu	Gln	Gln	Cys	Gln	Arg	His	
				500						505					510
Gly	Pro	Val	His	Cys	Ser	His	Gly	Gly	Gly	Arg	Phe	Leu	Ala	Gly	
				515						520					525

Val Ser Cys Met Asp Ser Ala Pro Asp Leu Val Met Asn Ala Gln		
530	535	540
Leu Val Gln Glu Thr Ala Tyr Leu Glu Asp Arg Pro Leu Ser Gln		
545	550	555
Leu Tyr Cys Ala His Glu Glu Asn Cys Leu Ser Lys Ser Ala Asp		
560	565	570
His Met Asp Trp Pro Tyr Gly Tyr Arg Arg Leu Leu Arg Phe Ser		
575	580	585
Thr Gln Ile Tyr Asn Leu Gly Arg Thr Asp Phe Arg Pro Lys Thr		
590	595	600
Gly Arg Asp Ser Trp Val Trp His Gln Cys His Arg His Tyr His		
605	610	615
Ser Ile Glu Val Phe Thr His Tyr Asp Leu Leu Thr Leu Asn Gly		
620	625	630
Ser Lys Val Ala Glu Gly His Lys Ala Ser Phe Cys Leu Glu Asp		
635	640	645
Thr Asn Cys Pro Thr Gly Leu Gln Arg Arg Tyr Ala Cys Ala Asn		
650	655	660
Phe Gly Glu Gln Gly Val Thr Val Gly Cys Trp Asp Thr Tyr Arg		
665	670	675
His Asp Ile Asp Cys Gln Trp Val Asp Ile Thr Asp Val Gly Pro		
680	685	690
Gly Asn Tyr Ile Phe Gln Val Ile Val Asn Pro His Tyr Glu Val		
695	700	705
Ala Glu Ser Asp Phe Ser Asn Asn Met Leu Gln Cys Arg Cys Lys		
710	715	720
Tyr Asp Gly His Arg Val Trp Leu His Asn Cys His Thr Gly Asn		
725	730	735
Ser Tyr Pro Ala Asn Ala Glu Leu Ser Leu Glu Gln Glu Gln Arg		
740	745	750
Leu Arg Asn Asn Leu Ile		
755		

&lt;210&gt; 2

&lt;211&gt; 544

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7472835CD1

&lt;400&gt; 2

Met Ala Lys Lys Ala Ile Ala Val Ile Gly Ala Gly Ile Ser Gly		
1	5	10
Leu Gly Ala Ile Lys Cys Cys Leu Asp Glu Asp Leu Glu Pro Thr		
20	25	30
Cys Phe Glu Arg Asn Asp Asp Ile Gly His Leu Trp Lys Phe Gln		
35	40	45
Lys Asn Thr Ser Glu Lys Met Pro Ser Ile Tyr Lys Ser Val Thr		
50	55	60
Ile Asn Thr Ser Lys Glu Met Met Cys Phe Ser Asp Phe Pro Val		
65	70	75
Pro Asp His Phe Pro Asn Tyr Met His Asn Ser Lys Leu Met Asp		
80	85	90
Tyr Phe Gly Met Tyr Ala Thr His Phe Gly Leu Leu Asn Tyr Ile		
95	100	105
Arg Phe Lys Thr Glu Val Gln Ser Val Arg Lys His Pro Asp Phe		
110	115	120
Ser Ile Asn Gly Gln Trp Asp Val Val Val Glu Thr Glu Glu Lys		
125	130	135
Gln Glu Thr Leu Val Phe Asp Gly Val Leu Val Cys Ser Gly His		

His Thr Asp Pro	140	145	150
Tyr Leu Pro Leu Gln	155	Ser Phe Pro Gly Met	Glu
Lys Phe Glu Gly Cys Tyr Phe His Ser	160	Arg Glu Tyr Lys Ser	Pro
Glu Asp Phe Ser Gly Lys Arg Ile Ile	170	Val Ile Gly Ile Gly Asn	180
Ser Gly Val Asp Ile Ala Val Glu Leu Ser	185	Arg Val Ala Lys Gln	195
Val Ile Phe Leu Ser Thr Arg Arg Gly	200	Ser Trp Ile Leu His Arg	210
Val Trp Asp Asn Gly Tyr Pro Met Asp	205	Ser Ser Phe Phe Thr Arg	225
Phe Asn Ser Phe Leu Gln Lys Ile Leu Thr	215	Thr Pro Gln Ile Asn	240
Asn Gln Leu Glu Lys Ile Met Asn Ser	230	Arg Phe Asn His Ala His	255
Cys Gly Leu Gln Pro Gln His Arg Ala	245	Leu Ser Gln His Pro Thr	270
Val Ser Asp Asp Leu Pro Asn His Ile	260	Ile Ser Gly Lys Val Gln	285
Val Lys Pro Ser Val Lys Glu Phe Thr	275	Glu Thr Asp Ala Ile Phe	300
Glu Asp Ser Thr Val Glu Glu Asn Ile	290	Asp Val Val Ile Phe Ala	315
Thr Gly Tyr Ser Phe Ser Phe Ser Phe	305	Leu Asp Gly Leu Ile Lys	330
Val Thr Asn Asn Glu Val Ser Leu Tyr	320	Lys Leu Met Phe Pro Pro	345
Asp Leu Glu Lys Pro Thr Leu Ala Val	335	Ile Gly Leu Ile Gln Pro	360
Leu Gly Ile Ile Leu Pro Ile Ala Glu	350	Leu Gln Ser Arg Trp Ala	375
Thr Arg Val Phe Lys Gly Leu Ile Lys	365	Leu Pro Ser Ala Glu Asn	390
Met Met Ala Asp Ile Ala Gln Arg Lys	380	Arg Ala Met Glu Lys Arg	405
Tyr Val Lys Thr Pro Arg His Thr Ile	395	Gln Val Asp His Ile Glu	420
Tyr Met Asp Glu Ile Ala Met Pro Ala	410	Gly Val Lys Pro Asn Leu	435
Leu Phe Leu Phe Leu Ser Asp Pro Lys	425	Leu Ala Met Glu Val Phe	450
Phe Gly Pro Cys Thr Pro Tyr Gln Tyr	440	His Leu His Gly Pro Glu	465
Lys Trp Asp Gly Ala Arg Arg Ala Asn	455	Leu Thr Gln Arg Glu Arg	480
Ile Ile Lys Pro Leu Arg Thr Arg Ile	470	Thr Ser Glu Asp Ser His	495
Pro Ser Ser Gln Leu Ser Trp Ile Lys	485	Met Ala Pro Val Ser Leu	510
Ala Phe Leu Ala Ala Gly Leu Ala Tyr	500	Phe Arg Tyr Thr Pro Tyr	525
Gly Lys Trp Lys	515		540
	530		
	535		

&lt;210&gt; 3

&lt;211&gt; 501

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7476203CD1

&lt;400&gt; 3

Met	Leu	Ser	Leu	Leu	Ser	Gly	Leu	Ala	Leu	Leu	Ala	Ile	Ser	Phe	1	5	10	15
Leu	Leu	Leu	Lys	Leu	Gly	Thr	Phe	Cys	Trp	Asp	Arg	Ser	Cys	Leu	20	25	30	35
Pro	Pro	Gly	Pro	Leu	Pro	Phe	Pro	Ile	Leu	Gly	Asn	Leu	Trp	Gln	40	45	50	55
Leu	Cys	Phe	Gln	Gln	Pro	His	Leu	Ser	Leu	Lys	Asn	Phe	Gln	Lys	60	65	70	75
Lys	Ile	Gly	Asn	Ile	Phe	Met	Asn	Leu	Gly	Ser	Ser	Val	Val	Pro	80	85	90	95
Leu	Ala	Leu	Pro	Leu	Leu	Pro	Val	Thr	Phe	His	Pro	Leu	Asn	Gln	100	105	110	115
Gly	Val	Leu	Cys	Lys	Pro	Leu	Ile	Thr	Phe	Pro	Lys	Pro	Phe	Pro	120	125	130	135
Thr	Arg	Asn	Pro	Gly	Ile	Ile	Cys	Ser	Ser	Gly	His	Thr	Trp	Arg	140	145	150	155
Gln	Lys	Arg	Arg	Phe	Cys	Leu	Val	Met	Ile	Arg	Gly	Leu	Gly	Leu	160	165	170	175
Gly	Lys	Leu	Ala	Leu	Glu	Val	Gln	Leu	Gln	Lys	Glu	Ala	Ala	Glu	180	185	190	195
Leu	Ala	Glu	Ala	Phe	Arg	Gln	Glu	Gln	Gly	Lys	Arg	Pro	Phe	Asp	200	205	210	215
Pro	Gln	Val	Ser	Ile	Val	Arg	Ser	Thr	Val	Arg	Val	Ile	Gly	Ala	220	225	230	235
Leu	Val	Phe	Gly	His	His	Phe	Leu	Leu	Glu	Asp	Pro	Ile	Phe	Gln	240	245	250	255
Glu	Leu	Thr	Gln	Ala	Ile	Asp	Phe	Gly	Leu	Ala	Phe	Val	Ser	Thr	260	265	270	275
Val	Trp	Arg	Gln	Leu	Tyr	Asp	Val	Phe	Pro	Trp	Ala	Leu	Cys	His	280	285	290	295
Leu	Pro	Gly	Pro	His	Gln	Glu	Ile	Phe	Arg	Tyr	Gln	Glu	Val	Val	300	305	310	315
Leu	Ser	Leu	Ile	His	Gln	Glu	Ile	Thr	Arg	His	Lys	Leu	Arg	Ala	320	325	330	335
Pro	Glu	Ala	Pro	Arg	Asp	Phe	Ile	Ser	Cys	Tyr	Leu	Ala	Gln	Ile	340	345	350	355
Ser	Lys	Ala	Met	Asp	Pro	Val	Ser	Thr	Phe	Asn	Gln	Glu	Asn		360	365	370	375
Leu	Val	Gln	Val	Val	Ile	Asp	Leu	Phe	Leu	Gly	Gly	Thr	Asp	Thr	380	385	390	395
Thr	Ala	Thr	Thr	Leu	Cys	Trp	Ala	Leu	Ile	His	Met	Ile	Gln	His	400	405	410	415
Gly	Ala	Val	Gln	Glu	Thr	Val	Gln	Leu	Glu	Leu	Asp	Glu	Val	Leu	420	425	430	435
Gly	Ala	Ala	Pro	Val	Val	Cys	Tyr	Glu	Asp	Arg	Lys	Arg	Leu	Pro				
Tyr	Thr	Met	Ala	Val	Leu	His	Asp	Val	Gln	Arg	Leu	Ser	Ser	Val				
Met	Ala	Met	Gly	Ala	Val	Arg	Gln	Cys	Val	Thr	Ser	Thr	Arg	Val				
Cys	Ser	Tyr	Pro	Val	Ser	Lys	Gly	Thr	Ile	Ile	Leu	Pro	Asn	Leu				
Ala	Ser	Val	Leu	Tyr	Asp	Pro	Glu	Cys	Trp	Glu	Thr	Pro	Arg	Gln				
Phe	Asn	Pro	Gly	His	Phe	Ser	Asp	Lys	Asp	Gly	Asn	Phe	Val	Ala				
Asn	Glu	Ala	Phe	Leu	Pro	Phe	Ser	Ala	Gly	Thr	Arg	Val	Tyr	Pro				

Ala Asp Gln Leu	Ala Gln Met Glu Leu	Phe Leu Met Phe Ala Thr
440		445
Leu Leu Arg Thr	Phe Arg Phe Gln Leu	Pro Glu Gly Ser Pro Gly
455		460
Leu Lys Leu Glu	Tyr Ile Phe Gly Gly	Thr Trp Gln Pro Gln Pro
470		475
Gln Glu Ile Cys	Ala Val Pro Arg Leu	Ser Ser Pro Ser Pro Gly
485		490
Pro Arg Glu Asp	Gly Leu	
500		

&lt;210&gt; 4

&lt;211&gt; 345

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7478583CD1

&lt;400&gt; 4

Met Lys Ala Ala	Val Trp Tyr Gly Gln	Lys Asp Val Arg Val Glu
1	5	10
Glu Arg Glu Pro	Lys Glu Leu Gln Asp	Asn Glu Val Lys Val Lys
20		25
Val Ser Trp Ala	Gly Ile Cys Gly Thr	Asp Leu His Glu Tyr Leu
35		40
Glu Gly Pro Ile	Phe Ile Ser Thr Glu	Lys Pro Asp Pro Phe Leu
50		55
Gly Gln Lys Ala	Pro Val Thr Leu Gly	His Glu Phe Ala Gly Val
65		70
Val Glu Glu Thr	Gly Ser Gln Val Thr	Lys Phe Asn Lys Gly Asp
80		85
Arg Val Val Val	Asn Pro Thr Val Ser	Lys Arg Glu Lys Glu Glu
95		100
Asn Ile Asp Leu	Tyr Asp Gly Tyr Ser	Phe Ile Gly Leu Gly Ser
110		115
Asp Gly Gly Phe	Ala Glu Phe Thr Asn	Ala Pro Glu Glu Asn Val
125		130
Tyr Lys Leu Pro	Asp Asn Val Ser Asp	Lys Glu Gly Ala Leu Val
140		145
Glu Pro Thr Ala	Val Ala Val Gln Ala	Ile Lys Glu Gly Glu Val
155		160
Leu Phe Gly Asp	Thr Val Ala Ile Phe	Gly Ala Gly Pro Ile Gly
170		175
Leu Leu Thr Val	Val Ala Ala Lys Ala	Ala Gly Ala Ser Lys Ile
185		190
Phe Val Phe Asp	Leu Ser Glu Glu Arg	Leu Ser Lys Ala Lys Ala
200		205
Leu Gly Ala Thr	His Ala Ile Asn Ser	Gly Lys Thr Asp Pro Val
215		220
Asp Val Ile Asn	Glu Tyr Thr Glu Asn	Gly Val Asp Val Ser Phe
230		235
Glu Val Ala Gly	Val Ala Pro Thr Leu	Lys Ser Ser Ile Asp Val
245		250
Thr Lys Ala Arg	Gly Thr Val Val Ile	Val Ser Ile Phe Gly His
260		265
Pro Ile Glu Trp	Asn Pro Met Gln Leu	Thr Asn Thr Gly Val Lys
275		280
Leu Thr Ser Thr	Ile Ala Tyr Thr Pro	Thr Thr Phe Gln Gln Thr
290		295
Ile Asp Leu Ile	Asn Glu Gly Asn Leu	Asn Val Lys Asp Val Val

Thr	Asp	Glu	Ile	305				310				315			
					Glu	Leu	Glu	Asn	Ile	Val	Glu	Ser	Gly	Phe	Glu
				320						325					330
Gln	Leu	Val	Asn	Asp	Lys	Ser	Gln	Ala	Lys	Ile	Leu	Ile	Lys	Leu	
				335						340					345

&lt;210&gt; 5

&lt;211&gt; 361

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7478585CD1

&lt;400&gt; 5

Met	Ser	Ala	Gln	Phe	Glu	Asn	Val	Gln	Asn	Pro	Ser	Ile	Thr	Arg	
1				5					10					15	
Glu	Asp	Val	Ala	Glu	Val	Leu	Val	Ser	Val	Leu	Thr	Asp	Glu	Thr	
				20					25					30	
Leu	Gln	Val	Val	Leu	Ala	Lys	Arg	Pro	Gln	Ser	Ile	Pro	Gln	Asp	
				35					40					45	
Asp	Val	Phe	Arg	Phe	Glu	Thr	Ile	Glu	Thr	Arg	Glu	Pro	His	Ala	
				50					55					60	
Gly	Glu	Val	Gln	Val	Glu	Ser	Ile	Tyr	Val	Ser	Val	Asp	Pro	Tyr	
				65					70					75	
Met	Arg	Gly	Arg	Met	Asn	Asp	Thr	Lys	Ser	Tyr	Val	Gln	Pro	Phe	
				80					85					90	
Gln	Val	Asn	Glu	Pro	Leu	Gln	Gly	His	Ile	Val	Gly	Lys	Val	Thr	
				95					100					105	
Gln	Ser	Asn	Asp	Glu	Arg	Leu	Ser	Val	Gly	Asp	Tyr	Val	Thr	Gly	
				110					115					120	
Ile	Leu	Pro	Trp	Lys	Lys	Ile	Asn	Thr	Val	Asn	Gly	Asp	Asp	Val	
				125					130					135	
Thr	Pro	Val	Pro	Ser	Lys	Asp	Val	Pro	Leu	His	Leu	Tyr	Leu	Ser	
				140					145					150	
Val	Leu	Gly	Met	Pro	Gly	Met	Thr	Ala	Tyr	Thr	Gly	Leu	Leu	Gln	
				155					160					165	
Ile	Gly	Gln	Pro	Gln	Ser	Gly	Glu	Thr	Val	Val	Val	Ser	Ala	Ala	
				170					175					180	
Ser	Gly	Ala	Val	Gly	Ser	Val	Val	Gly	Gln	Ile	Ala	Lys	Ile	Lys	
				185					190					195	
Gly	Ala	Lys	Val	Val	Gly	Ile	Ala	Gly	Gly	Lys	Gln	Lys	Thr	Thr	
				200					205					210	
Tyr	Leu	Thr	Asp	Glu	Leu	Gly	Phe	Asp	Ala	Ala	Ile	Asp	Tyr	Lys	
				215					220					225	
Gln	Asp	Asp	Phe	Ala	Gln	Gln	Leu	Glu	Ala	Ala	Val	Pro	Asp	Gly	
				230					235					240	
Ile	Asp	Val	Tyr	Phe	Glu	Asn	Val	Gly	Gly	Val	Ile	Ser	Asp	Glu	
				245					250					255	
Val	Phe	Lys	His	Leu	Asn	Arg	Phe	Ala	Arg	Val	Pro	Val	Cys	Gly	
				260					265					270	
Ala	Ile	Ser	Ala	Tyr	Asn	Asn	Glu	Lys	Asp	Asp	Ile	Gly	Pro	Arg	
				275					280					285	
Ile	Gln	Gly	Thr	Leu	Ile	Lys	Asn	Gln	Ala	Leu	Met	Gln	Gly	Phe	
				290					295					300	
Val	Val	Ala	Gln	Phe	Ala	Asp	His	Phe	Lys	Glu	Ala	Ser	Glu	Gln	
				305					310					315	
Leu	Ala	Gln	Trp	Val	Ser	Glu	Gly	Lys	Ile	Lys	Phe	Glu	Val	Thr	
				320					325					330	
Ile	Asp	Glu	Gly	Phe	Asp	Asn	Leu	Pro	Ser	Ala	Phe	Arg	Lys	Leu	

	335	340	345
Phe Thr Gly Asp Asn Phe Gly Lys Gln Val Val Lys Ile Lys Glu			
	350	355	360
Glu			

&lt;210&gt; 6

&lt;211&gt; 499

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7479904CD1

&lt;400&gt; 6

Met Asp Glu Lys Ser Asn Lys Leu Leu Leu Ala Leu Val Met Leu		
1	5	10
Phe Leu Phe Ala Val Ile Val Leu Gln Tyr Val Cys Pro Gly Thr		15
	20	25
Glu Cys Gln Leu Leu Arg Leu Gln Ala Phe Ser Ser Pro Val Pro		30
	35	40
Asp Pro Tyr Arg Ser Glu Asp Glu Ser Ser Ala Arg Phe Val Pro		45
	50	55
Arg Tyr Asn Phe Thr Arg Gly Asp Leu Leu Arg Lys Val Asp Phe		60
	65	70
Asp Ile Lys Gly Asp Asp Leu Ile Val Phe Leu His Ile Gln Lys		75
	80	85
Thr Gly Gly Thr Thr Phe Gly Arg His Leu Val Arg Asn Ile Gln		90
	95	100
Leu Glu Gln Pro Cys Glu Cys Arg Val Gly Gln Lys Lys Cys Thr		105
	110	115
Cys His Arg Pro Gly Lys Arg Glu Thr Trp Leu Phe Ser Arg Phe		120
	125	130
Ser Thr Gly Trp Ser Cys Gly Leu His Ala Asp Trp Thr Glu Leu		135
	140	145
Thr Ser Cys Val Pro Ser Val Val Asp Gly Lys Arg Asp Ala Arg		150
	155	160
Leu Arg Pro Ser Arg Trp Arg Ile Phe Gln Ile Leu Asp Ala Ala		165
	170	175
Ser Lys Asp Lys Arg Gly Ser Pro Asn Thr Asn Ala Gly Ala Asn		180
	185	190
Ser Pro Ser Ser Thr Lys Thr Arg Asn Thr Ser Lys Ser Gly Lys		195
	200	205
Asn Phe His Tyr Ile Thr Ile Leu Arg Asp Pro Val Ser Arg Tyr		210
	215	220
Leu Ser Glu Trp Arg His Val Gln Arg Gly Ala Thr Trp Lys Ala		225
	230	235
Ser Leu His Val Cys Asp Gly Arg Pro Pro Thr Ser Glu Glu Leu		240
	245	250
Pro Ser Cys Tyr Thr Gly Asp Asp Trp Ser Gly Cys Pro Leu Lys		255
	260	265
Glu Phe Met Asp Cys Pro Tyr Asn Leu Ala Asn Asn Arg Gln Val		270
	275	280
Arg Met Leu Ser Asp Leu Thr Leu Val Gly Cys Tyr Asn Leu Ser		285
	290	295
Val Met Pro Glu Lys Gln Arg Asn Lys Val Leu Leu Glu Ser Ala		300
	305	310
Lys Ser Asn Leu Lys His Met Ala Phe Phe Gly Leu Thr Glu Phe		315
	320	325
Gln Arg Lys Thr Gln Tyr Leu Phe Glu Lys Thr Phe Asn Met Asn		330
	335	340
		345

```

Phe Ile Ser Pro Phe Thr Gln Tyr Asn Thr Thr Arg Ala Ser Ser
350 355 360
Val Glu Ile Asn Glu Glu Ile Gln Lys Arg Ile Glu Gly Leu Asn
365 370 375
Phe Leu Asp Met Glu Leu Tyr Ser Tyr Ala Lys Asp Leu Phe Leu
380 385 390
Gln Arg Tyr Gln Phe Met Arg Gln Lys Glu His Gln Glu Ala Arg
395 400 405
Arg Lys Arg Gln Glu Gln Arg Lys Phe Leu Lys Gly Arg Leu Leu
410 415 420
Gln Thr His Phe Gln Ser Gln Gly Gln Gly Gln Ser Gln Asn Pro
425 430 435
Asn Gln Asn Gln Ser Gln Asn Pro Asn Pro Asn Ala Asn Gln Asn
440 445 450
Leu Thr Gln Asn Leu Met Gln Asn Leu Thr Gln Ser Leu Ser Gln
455 460 465
Lys Glu Asn Arg Glu Ser Pro Lys Gln Asn Ser Gly Lys Glu Gln
470 475 480
Asn Asp Asn Thr Ser Asn Gly Thr Asn Asp Tyr Ile Gly Ser Val
485 490 495
Glu Lys Trp Arg

```

```

<210> 7
<211> 222
<212> PRT
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7480367CD1

```

```

<400> 7
Met Ala Glu Lys Pro Lys Leu His Tyr Ser Asn Ala Arg Gly Ser
1 5 10 15
Met Glu Ser Ile Arg Trp Leu Leu Ala Ala Ala Gly Val Glu Leu
20 25 30
Glu Glu Lys Phe Leu Glu Ser Ala Glu Asp Leu Asp Lys Leu Arg
35 40 45
Asn Asp Gly Ser Leu Leu Phe Gln Gln Val Pro Met Val Glu Ile
50 55 60
Asp Gly Met Lys Leu Val Gln Thr Arg Ala Ile Leu Asn Tyr Ile
65 70 75
Ala Ser Lys Tyr Asn Leu Tyr Gly Lys Asp Met Lys Glu Arg Ala
80 85 90
Leu Ile Asp Met Tyr Thr Glu Gly Ile Val Asp Leu Thr Glu Met
95 100 105
Ile Leu Leu Leu Leu Ile Cys Gln Pro Glu Glu Arg Asp Ala Lys
110 115 120
Thr Ala Leu Val Lys Glu Lys Ile Lys Asn Arg Tyr Phe Pro Ala
125 130 135
Phe Glu Lys Val Leu Lys Ser His Arg Gln Asp Tyr Leu Val Gly
140 145 150
Asn Lys Leu Ser Trp Ala Asp Ile His Leu Val Glu Leu Phe Tyr
155 160 165
Tyr Val Glu Glu Leu Asp Ser Ser Leu Ile Ser Ser Phe Pro Leu
170 175 180
Leu Lys Ala Leu Lys Thr Arg Ile Ser Asn Leu Pro Thr Val Lys
185 190 195
Lys Phe Leu Gln Pro Gly Ser Gln Arg Lys Pro Pro Met Asp Glu
200 205 210
Lys Ser Leu Glu Glu Ala Arg Lys Ile Phe Arg Phe

```

215

220

<210> 8  
 <211> 330  
 <212> PRT  
 <213> Homo sapiens  
  
 <220>  
 <221> misc\_feature  
 <223> Incyte ID No: 8069390CD1

<400> 8  
 Met Ala Ala Val Asp Ser Phe Tyr Leu Leu Tyr Arg Glu Ile Ala  
 1 5 10 15  
 Arg Ser Cys Asn Cys Tyr Met Glu Ala Leu Ala Leu Val Gly Ala  
 20 25 30  
 Trp Tyr Thr Ala Arg Lys Ser Ile Thr Val Ile Cys Asp Phe Tyr  
 35 40 45  
 Ser Leu Ile Arg Leu His Phe Ile Pro Arg Leu Gly Ser Arg Ala  
 50 55 60  
 Asp Leu Ile Lys Gln Tyr Gly Arg Trp Ala Val Val Ser Gly Ala  
 65 70 75  
 Thr Asp Gly Ile Gly Lys Ala Tyr Ala Glu Leu Ala Ser Arg  
 80 85 90  
 Gly Leu Asn Ile Ile Leu Ile Ser Arg Asn Glu Glu Lys Leu Gln  
 95 100 105  
 Val Val Ala Lys Asp Ile Ala Asp Thr Tyr Lys Val Glu Thr Asp  
 110 115 120  
 Ile Ile Val Ala Asp Phe Ser Ser Gly Arg Glu Ile Tyr Leu Pro  
 125 130 135  
 Ile Arg Glu Ala Leu Lys Asp Lys Asp Val Gly Ile Leu Val Asn  
 140 145 150  
 Asn Val Gly Val Phe Tyr Pro Tyr Pro Gln Tyr Phe Thr Gln Leu  
 155 160 165  
 Ser Glu Asp Lys Leu Trp Asp Ile Ile Asn Val Asn Ile Ala Ala  
 170 175 180  
 Ala Ser Leu Met Val His Val Val Leu Pro Gly Met Val Glu Arg  
 185 190 195  
 Lys Lys Gly Ala Ile Val Thr Ile Ser Ser Gly Ser Cys Cys Lys  
 200 205 210  
 Pro Thr Pro Gln Leu Ala Ala Phe Ser Ala Ser Lys Ala Tyr Leu  
 215 220 225  
 Asp His Phe Ser Arg Ala Leu Gln Tyr Glu Tyr Ala Ser Lys Gly  
 230 235 240  
 Ile Phe Val Gln Ser Leu Ile Pro Phe Tyr Val Ala Thr Ser Met  
 245 250 255  
 Thr Ala Pro Ser Asn Phe Leu His Arg Cys Ser Trp Leu Val Pro  
 260 265 270  
 Ser Pro Lys Val Tyr Ala His His Ala Val Ser Thr Leu Gly Ile  
 275 280 285  
 Ser Lys Arg Thr Thr Gly Tyr Trp Ser His Ser Ile Gln Phe Leu  
 290 295 300  
 Phe Ala Gln Tyr Met Pro Glu Trp Leu Trp Val Trp Gly Ala Asn  
 305 310 315  
 Ile Leu Asn Arg Ser Leu Arg Lys Glu Ala Leu Ser Cys Thr Ala  
 320 325 330

<210> 9  
 <211> 303  
 <212> PRT  
 <213> Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7473869CD1

&lt;400&gt; 9

Met	Tyr	Val	Ser	Thr	Arg	Glu	Arg	Tyr	Lys	Trp	Leu	Arg	Phe	Ser
1				5					10					15
Glu	Asp	Cys	Leu	Tyr	Leu	Asn	Val	Tyr	Ala	Pro	Ala	Arg	Ala	Pro
			20						25					30
Gly	Asp	Pro	Gln	Leu	Pro	Val	Met	Val	Trp	Phe	Pro	Gly	Gly	Ala
			35						40					45
Phe	Ile	Val	Gly	Ala	Ala	Ser	Ser	Tyr	Glu	Gly	Ser	Asp	Leu	Ala
			50						55					60
Ala	Arg	Glu	Lys	Val	Val	Leu	Val	Phe	Leu	Gln	His	Arg	Leu	Gly
			65						70					75
Ile	Phe	Gly	Phe	Leu	Ser	Thr	Asp	Asp	Ser	His	Ala	Arg	Gly	Asn
			80						85					90
Trp	Gly	Leu	Leu	Asp	Gln	Met	Ala	Ala	Leu	Arg	Trp	Val	Gln	Glu
			95						100					105
Asn	Ile	Ala	Ala	Phe	Gly	Gly	Asp	Pro	Gly	Asn	Val	Thr	Leu	Phe
			110						115					120
Gly	Gln	Ser	Ala	Gly	Ala	Met	Ser	Ile	Ser	Gly	Leu	Met	Met	Ser
			125						130					135
Pro	Leu	Ala	Ser	Gly	Leu	Phe	His	Arg	Ala	Ile	Ser	Gln	Ser	Gly
			140						145					150
Thr	Ala	Leu	Phe	Arg	Leu	Phe	Ile	Thr	Ser	Asn	Pro	Leu	Lys	Val
			155						160					165
Ala	Lys	Lys	Val	Ala	His	Leu	Ala	Gly	Cys	Asn	His	Asn	Ser	Thr
			170						175					180
Gln	Ile	Leu	Val	Asn	Cys	Leu	Arg	Ala	Leu	Ser	Gly	Thr	Lys	Val
			185						190					195
Met	Arg	Val	Ser	Asn	Lys	Met	Arg	Phe	Leu	Gln	Leu	Asn	Phe	Gln
			200						205					210
Arg	Asp	Pro	Glu	Glu	Ile	Ile	Trp	Ser	Met	Ser	Pro	Val	Val	Asp
			215						220					225
Gly	Val	Val	Ile	Pro	Asp	Asp	Pro	Leu	Val	Leu	Leu	Thr	Gln	Gly
			230						235					240
Lys	Val	Ser	Ser	Val	Pro	Tyr	Leu	Leu	Gly	Val	Asn	Asn	Leu	Glu
			245						250					255
Phe	Asn	Trp	Leu	Leu	Pro	Tyr	Ile	Met	Lys	Phe	Pro	Leu	Asn	Arg
			260						265					270
Gln	Ala	Met	Arg	Lys	Glu	Thr	Ile	Thr	Lys	Met	Leu	Trp	Ser	Thr
			275						280					285
Arg	Thr	Leu	Leu	Val	Arg	Asp	Pro	Ala	Gly	Arg	Gly	Ala	Gln	Phe
			290						295					300
Gly	Gln	Gly												

&lt;210&gt; 10

&lt;211&gt; 584

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7478588CD1

&lt;400&gt; 10

Met	Pro	Ser	Thr	Val	Leu	Pro	Ser	Thr	Val	Leu	Pro	Ser	Leu	Leu
1				5					10					15
Pro	Thr	Ala	Gly	Ala	Gly	Trp	Ser	Met	Arg	Trp	Ile	Leu	Cys	Trp
			20						25					30

Ser	Leu	Thr	Leu	Cys	Leu	Met	Ala	Gln	Thr	Ala	Leu	Gly	Ala	Leu	
				35					40						45
His	Thr	Lys	Arg	Pro	Gln	Val	Val	Thr	Lys	Tyr	Gly	Thr	Leu	Gln	
				50					55						60
Gly	Lys	Gln	Met	His	Val	Gly	Lys	Thr	Pro	Ile	Gln	Val	Phe	Leu	
				65					70						75
Gly	Val	Pro	Phe	Ser	Arg	Pro	Pro	Leu	Gly	Ile	Leu	Arg	Phe	Ala	
				80					85						90
Pro	Pro	Glu	Pro	Pro	Glu	Pro	Trp	Lys	Gly	Ile	Arg	Asp	Ala	Thr	
				95					100						105
Thr	Tyr	Pro	Pro	Gly	Cys	Leu	Gln	Glu	Ser	Trp	Gly	Gln	Leu	Ala	
				110					115						120
Ser	Met	Tyr	Val	Ser	Thr	Arg	Glu	Arg	Tyr	Lys	Trp	Leu	Arg	Phe	
				125					130						135
Ser	Glu	Asp	Cys	Leu	Tyr	Leu	Asn	Val	Tyr	Ala	Pro	Ala	Arg	Ala	
				140					145						150
Pro	Gly	Asp	Pro	Gln	Leu	Pro	Val	Met	Val	Trp	Phe	Pro	Gly	Gly	
				155					160						165
Ala	Phe	Ile	Val	Gly	Ala	Ala	Ser	Ser	Tyr	Glu	Gly	Ser	Asp	Leu	
				170					175						180
Ala	Ala	Arg	Glu	Lys	Val	Val	Leu	Val	Phe	Leu	Gln	His	Arg	Leu	
				185					190						195
Gly	Ile	Phe	Gly	Phe	Leu	Ser	Thr	Asp	Asp	Ser	His	Ala	Arg	Gly	
				200					205						210
Asn	Trp	Gly	Leu	Leu	Asp	Gln	Met	Ala	Ala	Leu	Arg	Trp	Val	Gln	
				215					220						225
Glu	Asn	Ile	Ala	Ala	Phe	Gly	Gly	Asp	Pro	Gly	Asn	Val	Thr	Leu	
				230					235						240
Phe	Gly	Gln	Ser	Ala	Gly	Ala	Met	Ser	Ile	Ser	Gly	Leu	Met	Met	
				245					250						255
Ser	Pro	Leu	Ala	Ser	Gly	Leu	Phe	His	Arg	Ala	Ile	Ser	Gln	Ser	
				260					265						270
Gly	Thr	Ala	Leu	Phe	Arg	Leu	Phe	Ile	Thr	Ser	Asn	Pro	Leu	Lys	
				275					280						285
Val	Ala	Lys	Lys	Val	Ala	His	Leu	Ala	Gly	Cys	Asn	His	Asn	Ser	
				290					295						300
Thr	Gln	Ile	Leu	Val	Asn	Cys	Leu	Arg	Ala	Leu	Ser	Gly	Thr	Lys	
				305					310						315
Val	Met	Arg	Val	Ser	Asn	Lys	Met	Arg	Phe	Leu	Gln	Leu	Asn	Phe	
				320					325						330
Gln	Arg	Asp	Pro	Glu	Glu	Ile	Ile	Trp	Ser	Met	Ser	Pro	Val	Val	
				335					340						345
Asp	Gly	Val	Val	Ile	Pro	Asp	Asp	Pro	Leu	Val	Leu	Leu	Thr	Gln	
				350					355						360
Gly	Lys	Val	Ser	Ser	Val	Pro	Tyr	Leu	Leu	Gly	Val	Asn	Asn	Leu	
				365					370						375
Glu	Phe	Asn	Trp	Leu	Leu	Pro	Tyr	Ile	Met	Lys	Phe	Pro	Leu	Asn	
				380					385						390
Arg	Gln	Ala	Met	Arg	Lys	Glu	Thr	Ile	Thr	Lys	Met	Leu	Trp	Ser	
				395					400						405
Thr	Arg	Thr	Leu	Leu	Asn	Ile	Thr	Lys	Glu	Gln	Val	Pro	Leu	Val	
				410					415						420
Val	Glu	Glu	Tyr	Leu	Asp	Asn	Val	Asn	Glu	His	Asp	Trp	Lys	Met	
				425					430						435
Leu	Arg	Asn	Arg	Met	Met	Asp	Ile	Val	Gln	Asp	Ala	Thr	Phe	Val	
				440					445						450
Tyr	Ala	Thr	Leu	Gln	Thr	Ala	His	Tyr	His	Arg	Asp	Ala	Gly	Leu	
				455					460						465
Pro	Val	Tyr	Leu	Tyr	Glu	Phe	Glu	His	His	Ala	Arg	Gly	Ile	Ile	
				470					475						480
Val	Lys	Pro	Arg	Thr	Asp	Gly	Ala	Asp	His	Gly	Asp	Glu	Met	Tyr	
				485					490						495

Phe	Leu	Phe	Gly	Gly	Pro	Phe	Ala	Thr	Gly	Leu	Ser	Met	Gly	Lys	
				500					505						510
Glu	Lys	Ala	Leu	Ser	Leu	Gln	Met	Met	Lys	Tyr	Trp	Ala	Asn	Phe	
				515					520						525
Ala	Arg	Thr	Gly	Asn	Pro	Asn	Asp	Gly	Asn	Leu	Pro	Cys	Trp	Pro	
				530					535						540
Arg	Tyr	Asn	Lys	Asp	Glu	Lys	Tyr	Leu	Gln	Leu	Asp	Phe	Thr	Thr	
				545					550						555
Arg	Val	Gly	Met	Lys	Leu	Lys	Glu	Lys	Lys	Met	Ala	Phe	Trp	Met	
				560					565						570
Ser	Leu	Tyr	Gln	Ser	Gln	Arg	Pro	Glu	Lys	Gln	Arg	Gln	Phe		
				575					580						

&lt;210&gt; 11

&lt;211&gt; 508

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 55046125CD1

&lt;400&gt; 11

Met	His	Val	Leu	Arg	Arg	Arg	Trp	Asp	Leu	Gly	Ser	Leu	Cys	Arg	
1				5					10						15
Ala	Leu	Leu	Thr	Arg	Gly	Leu	Ala	Ala	Leu	Gly	His	Ser	Leu	Lys	
				20					25						30
His	Val	Leu	Gly	Ala	Ile	Phe	Ser	Lys	Ile	Phe	Gly	Pro	Met	Ala	
				35					40						45
Ser	Val	Gly	Asn	Met	Asp	Glu	Lys	Ser	Asn	Lys	Leu	Leu	Leu	Ala	
				50					55						60
Leu	Val	Met	Leu	Phe	Leu	Phe	Ala	Val	Ile	Val	Leu	Gln	Tyr	Val	
				65					70						75
Cys	Pro	Gly	Thr	Glu	Cys	Gln	Leu	Leu	Arg	Leu	Gln	Ala	Phe	Ser	
				80					85						90
Ser	Pro	Val	Pro	Asp	Pro	Tyr	Arg	Ser	Glu	Asp	Glu	Ser	Ser	Ala	
				95					100						105
Arg	Phe	Val	Pro	Arg	Tyr	Asn	Phe	Thr	Arg	Gly	Asp	Leu	Leu	Arg	
				110					115						120
Lys	Val	Asp	Phe	Asp	Ile	Lys	Gly	Asp	Asp	Leu	Ile	Val	Phe	Leu	
				125					130						135
His	Ile	Gln	Lys	Thr	Gly	Gly	Thr	Thr	Phe	Gly	Arg	His	Leu	Val	
				140					145						150
Arg	Asn	Ile	Gln	Leu	Glu	Gln	Pro	Cys	Glu	Cys	Arg	Val	Gly	Gln	
				155					160						165
Lys	Lys	Cys	Thr	Cys	His	Arg	Pro	Gly	Lys	Arg	Glu	Thr	Trp	Leu	
				170					175						180
Phe	Ser	Arg	Phe	Ser	Thr	Gly	Trp	Ser	Cys	Gly	Leu	His	Ala	Asp	
				185					190						195
Trp	Thr	Glu	Leu	Thr	Ser	Cys	Val	Pro	Ser	Val	Val	Asp	Gly	Lys	
				200					205						210
Arg	Asp	Ala	Arg	Leu	Arg	Pro	Ser	Arg	Asn	Phe	His	Tyr	Ile	Thr	
				215					220						225
Ile	Leu	Arg	Asp	Pro	Val	Ser	Arg	Tyr	Leu	Ser	Glu	Trp	Arg	His	
				230					235						240
Val	Gln	Arg	Gly	Ala	Thr	Trp	Lys	Ala	Ser	Leu	His	Val	Cys	Asp	
				245					250						255
Gly	Arg	Pro	Pro	Thr	Ser	Glu	Glu	Leu	Pro	Ser	Cys	Tyr	Thr	Gly	
				260					265						270
Asp	Asp	Trp	Ser	Gly	Cys	Pro	Leu	Lys	Glu	Phe	Met	Asp	Cys	Pro	
				275					280						285
Tyr	Asn	Leu	Ala	Asn	Asn	Arg	Gln	Val	Arg	Met	Leu	Ser	Asp	Leu	

Thr Leu Val Gly Cys Tyr Asn Leu Ser Val Met Pro Glu Lys Gln	290	295	300
305	310	315	
Arg Asn Lys Val Leu Leu Glu Ser Ala Lys Ser Asn Leu Lys His	320	325	330
Met Ala Phe Phe Gly Leu Thr Glu Phe Gln Arg Lys Thr Gln Tyr	335	340	345
Leu Phe Glu Lys Thr Phe Asn Met Asn Phe Ile Ser Pro Phe Thr	350	355	360
Gln Tyr Asn Thr Thr Arg Ala Ser Ser Val Glu Ile Asn Glu Glu	365	370	375
Ile Gln Lys Arg Ile Glu Gly Leu Asn Phe Leu Asp Met Glu Leu	380	385	390
Tyr Ser Tyr Ala Lys Asp Leu Phe Leu Gln Arg Tyr Gln Phe Met	395	400	405
Arg Gln Lys Glu His Gln Glu Ala Arg Arg Lys Arg Gln Glu Gln	410	415	420
Arg Lys Phe Leu Lys Gly Arg Leu Leu Gln Thr His Phe Gln Ser	425	430	435
Gln Gly Gln Gly Gln Ser Gln Asn Pro Asn Gln Asn Gln Ser Gln	440	445	450
Asn Pro Asn Pro Asn Ala Asn Gln Asn Leu Thr Gln Asn Leu Met	455	460	465
Gln Asn Leu Thr Gln Ser Leu Ser Gln Lys Glu Asn Arg Glu Ser	470	475	480
Pro Lys Gln Asn Ser Gly Lys Glu Gln Asn Asp Asn Thr Ser Asn	485	490	495
Gly Thr Asn Asp Tyr Ile Gly Ser Val Glu Lys Trp Arg	500	505	

&lt;210&gt; 12

&lt;211&gt; 439

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 3538709CD1

&lt;400&gt; 12

Met Leu Thr Gly Val Thr Asp Gly Ile Phe Cys Cys Leu Leu Gly	1	5	10	15
Thr Pro Pro Asn Ala Val Gly Pro Leu Glu Ser Val Glu Ser Ser	20	25	30	
Asp Gly Tyr Thr Phe Val Glu Val Lys Pro Gly Arg Val Leu Arg	35	40	45	
Val Lys His Ala Gly Pro Ala Pro Ala Ala Ala Pro Pro Pro Pro	50	55	60	
Ser Ser Ala Ser Ser Asp Ala Ala Gln Gly Asp Leu Ser Gly Leu	65	70	75	
Val Arg Cys Gln Arg Arg Ile Thr Val Tyr Arg Asn Gly Arg Leu	80	85	90	
Leu Val Glu Asn Leu Gly Arg Ala Pro Arg Ala Asp Leu Leu His	95	100	105	
Gly Gln Asn Gly Ser Gly Glu Pro Pro Ala Ala Leu Glu Val Glu	110	115	120	
Leu Ala Asp Pro Ala Gly Ser Asp Gly Arg Leu Ala Pro Gly Ser	125	130	135	
Ala Gly Ser Gly Ser Gly Ser Gly Ser Gly Gly Arg Arg Arg Arg	140	145	150	
Ala Arg Arg Pro Lys Arg Thr Ile His Ile Asp Cys Glu Lys Arg	155	160	165	

```

Ile Thr Ser Cys Lys Gly Ala Gln Ala Asp Val Val Leu Phe Phe
170 175 180
Ile His Gly Val Gly Gly Ser Leu Ala Ile Trp Lys Glu Gln Leu
185 190 195
Asp Phe Phe Val Arg Leu Gly Tyr Glu Val Val Ala Pro Asp Leu
200 205 210
Ala Gly His Gly Ala Ser Ser Ala Pro Gln Val Ala Ala Ala Tyr
215 220 225
Thr Phe Tyr Ala Leu Ala Glu Asp Met Arg Ala Ile Phe Lys Arg
230 235 240
Tyr Ala Lys Lys Arg Asn Val Leu Ile Gly His Ser Tyr Gly Val
245 250 255
Ser Phe Cys Thr Phe Leu Ala His Glu Tyr Pro Asp Leu Val His
260 265 270
Lys Val Ile Met Ile Asn Gly Gly Gly Pro Thr Ala Leu Glu Pro
275 280 285
Ser Phe Cys Ser Ile Phe Asn Met Pro Thr Cys Val Leu His Cys
290 295 300
Leu Ser Pro Cys Leu Ala Trp Ser Phe Leu Lys Ala Gly Phe Ala
305 310 315
Arg Gln Gly Ala Lys Glu Lys Gln Leu Leu Lys Glu Gly Asn Ala
320 325 330
Phe Asn Val Ser Ser Phe Val Leu Arg Ala Met Met Ser Gly Gln
335 340 345
Tyr Trp Pro Glu Gly Asp Glu Val Tyr His Ala Glu Leu Thr Val
350 355 360
Pro Val Leu Leu Val His Gly Met His Asp Lys Phe Val Pro Val
365 370 375
Glu Glu Asp Gln Arg Met Ala Glu Ile Leu Leu Leu Ala Phe Leu
380 385 390
Lys Leu Ile Asp Glu Gly Ser His Met Val Met Leu Glu Cys Pro
395 400 405
Glu Thr Val Asn Thr Leu Leu His Glu Phe Leu Leu Trp Glu Pro
410 415 420
Glu Pro Ser Pro Lys Ala Leu Pro Glu Pro Leu Pro Ala Pro Pro
425 430 435
Glu Asp Lys Lys

```

&lt;210&gt; 13

&lt;211&gt; 514

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 71563101CD1

&lt;400&gt; 13

```

Met Thr Leu Ile Trp Arg His Leu Leu Arg Pro Leu Cys Leu Val
1 5 10 15
Thr Ser Ala Pro Arg Ile Leu Glu Met His Pro Phe Leu Ser Leu
20 25 30
Gly Thr Ser Arg Thr Ser Val Thr Lys Leu Ser Leu His Thr Lys
35 40 45
Pro Arg Met Pro Pro Cys Asp Phe Met Pro Glu Arg Tyr Gln Ser
50 55 60
Leu Gly Tyr Asn Arg Val Leu Glu Ile His Lys Glu His Leu Ser
65 70 75
Pro Val Val Thr Ala Tyr Phe Gln Lys Pro Leu Leu Leu His Gln
80 85 90
Gly His Met Glu Trp Leu Phe Asp Ala Glu Gly Asn Arg Tyr Leu

```

	95		100		105
Asp Phe Phe Ser	Gly Ile Val Thr Val	Ser Val Gly His Cys His			
	110		115		120
Pro Lys Val Asn	Ala Val Ala Gln Lys	Gln Leu Gly Arg Leu Trp			
	125		130		135
His Thr Ser Thr	Val Phe Phe His Pro	Pro Met His Glu Tyr Ala			
	140		145		150
Glu Lys Leu Ala	Ala Leu Leu Pro Glu	Pro Leu Lys Val Ile Phe			
	155		160		165
Leu Val Asn Ser	Gly Ser Glu Ala Asn	Glu Leu Ala Met Leu Met			
	170		175		180
Ala Arg Ala His	Ser Asn Asn Ile Asp	Ile Ile Ser Phe Arg Gly			
	185		190		195
Ala Tyr His Gly	Cys Ser Pro Tyr Thr	Leu Gly Leu Thr Asn Val			
	200		205		210
Gly Ile Tyr Lys	Met Glu Leu Pro Gly	Gly Thr Gly Cys Gln Pro			
	215		220		225
Thr Met Cys Pro	Asp Val Phe Arg Gly	Pro Trp Gly Gly Ser His			
	230		235		240
Cys Arg Asp Ser	Pro Val Gln Thr Ile	Arg Lys Cys Ser Cys Ala			
	245		250		255
Pro Asp Cys Cys	Gln Ala Lys Asp Gln	Tyr Ile Glu Gln Phe Lys			
	260		265		270
Asp Thr Leu Ser	Thr Ser Val Ala Lys	Ser Ile Ala Gly Phe Phe			
	275		280		285
Ala Glu Pro Ile	Gln Gly Val Asn Gly	Val Val Gln Tyr Pro Lys			
	290		295		300
Gly Phe Leu Lys	Glu Ala Phe Glu Leu	Val Arg Ala Arg Gly Gly			
	305		310		315
Val Cys Ile Ala	Asp Glu Val Gln Thr	Gly Phe Gly Arg Leu Gly			
	320		325		330
Ser His Phe Trp	Gly Phe Gln Thr His	Asp Val Leu Pro Asp Ile			
	335		340		345
Val Thr Met Ala	Lys Gly Ile Gly Asn	Gly Phe Pro Met Ala Ala			
	350		355		360
Val Ile Thr Thr	Pro Glu Ile Ala Lys	Ser Leu Ala Lys Cys Leu			
	365		370		375
Gln His Phe Asn	Thr Phe Gly Gly Asn	Pro Met Ala Cys Ala Ile			
	380		385		390
Gly Ser Ala Val	Leu Glu Val Ile Lys	Glu Glu Asn Leu Gln Glu			
	395		400		405
Asn Ser Gln Glu	Val Gly Thr Tyr Met	Leu Leu Lys Phe Ala Lys			
	410		415		420
Leu Arg Asp Glu	Phe Glu Ile Val Gly	Asp Val Arg Gly Lys Gly			
	425		430		435
Leu Met Ile Gly	Ile Glu Met Val Gln	Asp Lys Ile Ser Cys Arg			
	440		445		450
Pro Leu Pro Arg	Glu Glu Val Asn Gln	Ile His Glu Asp Cys Lys			
	455		460		465
His Met Gly Leu	Leu Val Gly Arg Gly	Ser Ile Phe Ser Gln Thr			
	470		475		480
Phe Arg Ile Ala	Pro Ser Met Cys Ile	Thr Lys Pro Glu Val Asp			
	485		490		495
Phe Ala Val Glu	Val Phe Arg Ser Ala	Leu Thr Gln His Met Glu			
	500		505		510
Arg Arg Ala Lys					

<210> 14  
 <211> 226  
 <212> PRT  
 <213> Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7472027CD1

&lt;400&gt; 14

```

Met Arg Leu Cys Glu Lys Thr Glu Leu Gln Leu Ile Gly Val Pro
 1           5           10           15
Glu Ser Asp Arg Glu Asn Gly Thr Lys Leu Glu Asn Thr Phe Gln
          20           25           30
Asp Ile Ile Gln Glu Asn Phe Pro Asn Leu Ala Arg Gln Ala Asn
          35           40           45
Ile Gln Ile Gln Met Ala Gly Gly Ser Ile Trp Ile Glu Gly Ile
          50           55           60
Pro Phe Pro Ser Asn Asn Phe Thr Asp Leu Arg Arg Leu Gln Asp
          65           70           75
Glu Ile Val Leu Arg Asp Glu Asp Val Ile Thr Leu Ser Tyr Pro
          80           85           90
Lys Ser Gly Ser Phe Trp Ile Val Glu Ile Ile Ser Leu Ile His
          95          100          105
Ser Lys Gly Asp Pro Ser Trp Val Gln Ser Val Val Pro Trp Asp
          110          115          120
Arg Ser Pro Trp Ile Glu Val Lys Arg Lys Lys Ala Gly Leu Glu
          125          130          135
Ser Gln Lys Gly Pro His Leu Tyr Thr Ser His Leu Pro Ile Gln
          140          145          150
Leu Phe Pro Lys Ser Phe Leu Asn Ser Lys Ala Lys Cys Ile Tyr
          155          160          165
Pro His Val Leu Met Leu Val Val Leu Ile Leu Gly His Lys Ser
          170          175          180
Gln Trp Ser Ile Ala Ile Lys Ile Ser Glu Asn Ala Glu Ala Thr
          185          190          195
Ser Lys Leu Gly Asn Gly Gln Arg Leu Glu Glu Phe Gly Gly Leu
          200          205          210
Arg Arg Arg Gln Glu Asp Glu Arg Ser Leu Glu Phe Leu Arg Asp
          215          220          225
Cys

```

&lt;210&gt; 15

&lt;211&gt; 121

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7480358CD1

&lt;400&gt; 15

```

Met Met Lys Val Met Tyr Met Leu Lys Gly Gln Ser Pro Val Gln
 1           5           10           15
Gly Thr Ile His Phe Glu Gln Lys Glu Asn Glu Pro Phe Met Val
          20           25           30
Ser Glu Cys Ile Thr Gly Leu Thr Glu Arg Gln His Arg Phe His
          35           40           45
Val His Gln Phe Gly Asp Asn Thr Pro Gly Cys Thr Arg Ala Val
          50           55           60
Pro Tyr Phe Asn Pro Leu Thr Lys Asn His Ser Gly Pro Arg Ile
          65           70           75
Lys Arg Gly Arg Glu Glu Thr Trp Val Met Trp Pro Leu Ala Lys
          80           85           90
Met Cys Arg His Met Ser Val Glu Asp Ser Leu Val Ser Leu Ser
          95          100          105

```

Gly His Tyr Ser Ile Thr Ala His Thr Met Val Ser Met Thr Thr  
 110 115 120

Arg

<210> 16

<211> 486

<212> PRT

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 1618256CD1

<400> 16

Met Gly Pro Leu Ser Pro Ala Arg Thr Leu Arg Leu Trp Gly Pro  
 1 5 10 15  
 Arg Ser Leu Gly Val Ala Leu Gly Val Phe Met Thr Ile Gly Phe  
 20 25 30  
 Ala Leu Gln Leu Leu Gly Gly Pro Phe Gln Arg Arg Leu Pro Gly  
 35 40 45  
 Leu Gln Leu Arg Gln Pro Ser Ala Pro Ser Leu Arg Pro Ala Leu  
 50 55 60  
 Pro Ser Cys Pro Pro Arg Gln Arg Leu Val Phe Leu Lys Thr His  
 65 70 75  
 Lys Ser Gly Ser Ser Ser Val Leu Ser Leu Leu His Arg Tyr Gly  
 80 85 90  
 Asp Gln His Gly Leu Arg Phe Ala Leu Pro Ala Arg Tyr Gln Phe  
 95 100 105  
 Gly Tyr Pro Lys Leu Phe Gln Ala Ser Arg Val Lys Gly Tyr Arg  
 110 115 120  
 Pro Gln Gly Gly Gly Thr Gln Leu Pro Phe His Ile Leu Cys His  
 125 130 135  
 His Met Arg Phe Asn Leu Lys Glu Val Leu Gln Val Met Pro Ser  
 140 145 150  
 Asp Ser Phe Phe Phe Ser Ile Val Arg Asp Pro Ala Ala Leu Ala  
 155 160 165  
 Arg Ser Ala Phe Ser Tyr Tyr Lys Ser Thr Ser Ser Ala Phe Arg  
 170 175 180  
 Lys Ser Pro Ser Ser Leu Ala Ala Phe Leu Ala Asn Pro Arg Gly Phe  
 185 190 195  
 Tyr Arg Pro Gly Ala Arg Gly Asp His Tyr Ala Arg Asn Leu Leu  
 200 205 210  
 Trp Phe Asp Phe Gly Leu Pro Phe Pro Pro Glu Lys Arg Ala Lys  
 215 220 225  
 Arg Gly Asn Ile His Pro Pro Arg Asp Pro Asn Pro Pro Gln Leu  
 230 235 240  
 Gln Val Leu Pro Ser Gly Ala Gly Pro Arg Ala Gln Thr Leu Asn  
 245 250 255  
 Pro Asn Ala Leu Ile His Pro Val Ser Thr Val Thr Asp His Arg  
 260 265 270  
 Ser Gln Ile Ser Ser Pro Ala Ser Phe Asp Leu Gly Ser Ser Ser  
 275 280 285  
 Phe Ile Gln Trp Gly Leu Ala Trp Leu Asp Ser Val Phe Asp Leu  
 290 295 300  
 Val Met Val Ala Glu Tyr Phe Asp Glu Ser Leu Val Leu Leu Ala  
 305 310 315  
 Asp Ala Leu Cys Trp Gly Leu Asp Asp Val Val Gly Phe Met His  
 320 325 330  
 Asn Ala Gln Ala Gly His Lys Gln Gly Leu Ser Thr Val Ser Asn  
 335 340 345  
 Ser Gly Leu Thr Ala Glu Asp Arg Gln Leu Thr Ala Arg Ala Arg

Ala Trp Asn Asn	350	355	360
Leu Asp Trp Ala Leu Tyr Val His Phe Asn Arg			
	365	370	375
Ser Leu Trp Ala Arg Ile Glu Lys Tyr Gly Gln Gly Arg Leu Gln			
	380	385	390
Thr Ala Val Ala Glu Leu Arg Ala Arg Arg Glu Ala Leu Ala Lys			
	395	400	405
His Cys Leu Val Gly Gly Glu Ala Ser Asp Pro Lys Tyr Ile Thr			
	410	415	420
Asp Arg Arg Phe Arg Pro Phe Gln Phe Gly Ser Ala Lys Val Leu			
	425	430	435
Gly Tyr Ile Leu Arg Ser Gly Leu Ser Pro Gln Asp Gln Glu Glu			
	440	445	450
Cys Glu Arg Leu Ala Thr Pro Glu Leu Gln Tyr Lys Asp Lys Leu			
	455	460	465
Asp Val Lys Gln Phe Pro Pro Thr Val Ser Leu Pro Leu Lys Thr			
	470	475	480
Ser Arg Pro Leu Ser Pro			
	485		

&lt;210&gt; 17

&lt;211&gt; 649

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 3387823CD1

&lt;400&gt; 17

Met Tyr Ile Ser Cys Leu Ser Leu Ser Leu Phe Phe Leu Ser Gly			
1	5	10	15
Pro Leu Gln Arg Val Leu Glu Val Ser Asn His Trp Trp Tyr Ser			
	20	25	30
Met Leu Ile Leu Pro Pro Leu Leu Lys Asp Ser Val Ala Ala Pro			
	35	40	45
Leu Leu Ser Ala Tyr Tyr Pro Asp Cys Val Gly Met Ser Pro Ser			
	50	55	60
Cys Thr Ser Thr Asn Arg Ala Ala Ala Thr Gly Asn Ala Ser Pro			
	65	70	75
Gly Lys Leu Glu His Ser Lys Ala Ala Leu Ser Val His Val Pro			
	80	85	90
Gly Met Asn Arg Tyr Phe Gln Pro Phe Tyr Gln Pro Asn Glu Cys			
	95	100	105
Gly Lys Ala Leu Cys Val Arg Pro Asp Val Met Glu Leu Asp Glu			
	110	115	120
Leu Tyr Glu Phe Pro Glu Tyr Ser Arg Asp Pro Thr Met Tyr Leu			
	125	130	135
Ala Leu Arg Asn Leu Ile Leu Ala Leu Trp Tyr Thr Asn Cys Lys			
	140	145	150
Glu Ala Leu Thr Pro Gln Lys Cys Ile Pro His Ile Ile Val Arg			
	155	160	165
Gly Leu Val Arg Ile Arg Cys Val Gln Glu Val Glu Arg Ile Leu			
	170	175	180
Tyr Phe Met Thr Arg Lys Gly Leu Ile Asn Thr Gly Val Leu Ser			
	185	190	195
Val Gly Ala Asp Gln Tyr Leu Leu Pro Lys Asp Tyr His Asn Lys			
	200	205	210
Ser Val Ile Ile Ile Gly Ala Gly Pro Ala Gly Leu Ala Ala Ala			
	215	220	225
Arg Gln Leu His Asn Phe Gly Ile Lys Val Thr Val Leu Glu Ala			
	230	235	240

Lys	Asp	Arg	Ile	Gly	Gly	Arg	Val	Trp	Asp	Asp	Lys	Ser	Phe	Lys
				245					250					255
Gly	Val	Thr	Val	Gly	Arg	Gly	Ala	Gln	Ile	Val	Asn	Gly	Cys	Ile
				260					265					270
Asn	Asn	Pro	Val	Ala	Leu	Met	Cys	Glu	Gln	Leu	Gly	Ile	Ser	Met
				275					280					285
His	Lys	Phe	Gly	Glu	Arg	Cys	Asp	Leu	Ile	Gln	Glu	Gly	Gly	Arg
				290					295					300
Ile	Thr	Asp	Pro	Thr	Ile	Asp	Lys	Arg	Met	Asp	Phe	His	Phe	Asn
				305					310					315
Ala	Leu	Leu	Asp	Val	Val	Ser	Glu	Trp	Arg	Lys	Asp	Lys	Thr	Gln
				320					325					330
Leu	Gln	Asp	Val	Pro	Leu	Gly	Glu	Lys	Ile	Glu	Glu	Ile	Tyr	Lys
				335					340					345
Ala	Phe	Ile	Lys	Glu	Ser	Gly	Ile	Gln	Phe	Ser	Glu	Leu	Glu	Gly
				350					355					360
Gln	Val	Leu	Gln	Phe	His	Leu	Ser	Asn	Leu	Glu	Tyr	Ala	Cys	Gly
				365					370					375
Ser	Asn	Leu	His	Gln	Val	Ser	Ala	Arg	Ser	Trp	Asp	His	Asn	Glu
				380					385					390
Phe	Phe	Ala	Gln	Phe	Ala	Gly	Asp	His	Thr	Leu	Leu	Thr	Pro	Gly
				395					400					405
Tyr	Ser	Val	Ile	Ile	Glu	Lys	Leu	Ala	Glu	Gly	Leu	Asp	Ile	Gln
				410					415					420
Leu	Lys	Ser	Pro	Val	Gln	Cys	Ile	Asp	Tyr	Ser	Gly	Asp	Glu	Val
				425					430					435
Gln	Val	Thr	Thr	Thr	Asp	Gly	Thr	Gly	Tyr	Ser	Ala	Gln	Lys	Val
				440					445					450
Leu	Val	Thr	Val	Pro	Leu	Ala	Leu	Leu	Gln	Lys	Gly	Ala	Ile	Gln
				455					460					465
Phe	Asn	Pro	Pro	Leu	Ser	Glu	Lys	Lys	Met	Lys	Ala	Ile	Asn	Ser
				470					475					480
Leu	Gly	Ala	Gly	Ile	Ile	Glu	Lys	Ile	Ala	Leu	Gln	Phe	Pro	Tyr
				485					490					495
Arg	Phe	Trp	Asp	Ser	Lys	Val	Gln	Gly	Ala	Asp	Phe	Phe	Gly	His
				500					505					510
Val	Pro	Pro	Ser	Ala	Ser	Lys	Arg	Gly	Leu	Phe	Ala	Val	Phe	Tyr
				515					520					525
Asp	Met	Asp	Pro	Gln	Lys	Lys	His	Ser	Val	Leu	Met	Ser	Val	Ile
				530					535					540
Ala	Gly	Glu	Ala	Val	Ala	Ser	Val	Arg	Thr	Leu	Asp	Asp	Lys	Gln
				545					550					555
Val	Leu	Gln	Gln	Cys	Met	Ala	Thr	Leu	Arg	Glu	Leu	Phe	Lys	Glu
				560					565					570
Gln	Glu	Val	Pro	Asp	Pro	Thr	Lys	Tyr	Phe	Val	Thr	Arg	Trp	Ser
				575					580					585
Thr	Asp	Pro	Trp	Ile	Gln	Met	Ala	Tyr	Ser	Phe	Val	Lys	Thr	Gly
				590					595					600
Gly	Ser	Gly	Glu	Ala	Tyr	Asp	Ile	Ile	Ala	Glu	Asp	Ile	Gln	Gly
				605					610					615
Thr	Val	Phe	Phe	Ala	Gly	Glu	Ala	Thr	Asn	Arg	His	Phe	Pro	Gln
				620					625					630
Thr	Val	Thr	Gly	Ala	Tyr	Leu	Ser	Gly	Val	Arg	Glu	Ala	Ser	Lys
				635					640					645
Ile	Ala	Ala	Phe											

&lt;210&gt; 18

&lt;211&gt; 258

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 55142051CD1

&lt;400&gt; 18

```

Met Ser Pro Ala Ile Ala Leu Ala Phe Leu Pro Leu Val Val Thr
 1           5           10           15
Leu Leu Val Arg Tyr Arg His Tyr Phe Arg Leu Leu Val Arg Thr
           20           25           30
Val Leu Leu Arg Ser Leu Arg Asp Cys Leu Ser Gly Leu Arg Ile
           35           40           45
Glu Glu Arg Ala Phe Ser Tyr Val Leu Thr His Ala Leu Pro Gly
           50           55           60
Asp Pro Gly His Ile Leu Thr Thr Leu Asp His Trp Ser Ser Arg
           65           70           75
Cys Glu Tyr Leu Ser His Met Gly Pro Val Lys Gly Gln Ile Leu
           80           85           90
Met Arg Leu Val Glu Glu Lys Ala Pro Ala Cys Val Leu Glu Leu
           95          100          105
Gly Thr Tyr Cys Gly Tyr Ser Thr Leu Leu Ile Ala Arg Ala Leu
          110          115          120
Pro Pro Gly Gly Arg Leu Leu Thr Val Glu Arg Asp Pro Arg Thr
          125          130          135
Ala Ala Val Ala Glu Lys Leu Ile Arg Leu Ala Gly Phe Asp Glu
          140          145          150
His Met Val Glu Leu Ile Val Gly Ser Ser Glu Asp Val Ile Pro
          155          160          165
Cys Leu Arg Thr Gln Tyr Gln Leu Ser Arg Ala Asp Leu Val Leu
          170          175          180
Leu Ala His Arg Pro Arg Cys Tyr Leu Arg Asp Leu Gln Leu Leu
          185          190          195
Glu Ala His Ala Leu Leu Pro Ala Gly Ala Thr Val Leu Ala Asp
          200          205          210
His Val Leu Phe Pro Gly Ala Pro Arg Phe Leu Gln Tyr Ala Lys
          215          220          225
Ser Cys Gly Arg Tyr Arg Cys Arg Leu His His Thr Gly Leu Pro
          230          235          240
Asp Phe Pro Ala Ile Lys Asp Gly Ile Ala Gln Leu Thr Tyr Ala
          245          250          255
Gly Pro Gly

```

&lt;210&gt; 19

&lt;211&gt; 544

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7395274CD1

&lt;400&gt; 19

```

Met Ser Ser Pro Gly Pro Ser Gln Pro Pro Ala Glu Asp Pro Pro
 1           5           10           15
Trp Pro Ala Arg Leu Leu Arg Ala Pro Leu Gly Leu Leu Arg Leu
           20           25           30
Asp Pro Ser Gly Gly Ala Leu Leu Leu Cys Gly Leu Val Ala Leu
           35           40           45
Leu Gly Trp Ser Trp Leu Arg Arg Arg Arg Ala Arg Gly Ile Pro
           50           55           60
Pro Gly Pro Thr Pro Trp Pro Leu Val Gly Asn Phe Gly His Val
           65           70           75

```

Leu	Leu	Pro	Pro	Phe	Leu	Arg	Arg	Arg	Ser	Trp	Leu	Ser	Ser	Arg	80	85	90
Thr	Arg	Ala	Ala	Gly	Ile	Asp	Pro	Ser	Val	Ile	Gly	Pro	Gln	Val	95	100	105
Leu	Leu	Ala	His	Leu	Ala	Arg	Val	Tyr	Gly	Ser	Ile	Phe	Ser	Phe	110	115	120
Phe	Ile	Gly	His	Tyr	Leu	Val	Val	Val	Leu	Ser	Asp	Phe	His	Ser	125	130	135
Val	Arg	Glu	Ala	Leu	Val	Gln	Gln	Ala	Glu	Val	Phe	Ser	Asp	Arg	140	145	150
Pro	Arg	Val	Pro	Leu	Ile	Ser	Ile	Val	Thr	Lys	Glu	Lys	Gly	Val	155	160	165
Val	Phe	Ala	His	Tyr	Gly	Pro	Val	Trp	Arg	Gln	Gln	Arg	Lys	Phe	170	175	180
Ser	His	Ser	Thr	Leu	Arg	His	Phe	Gly	Leu	Gly	Lys	Leu	Ser	Leu	185	190	195
Glu	Pro	Lys	Ile	Ile	Glu	Glu	Phe	Lys	Tyr	Val	Lys	Ala	Glu	Met	200	205	210
Gln	Lys	His	Gly	Glu	Asp	Pro	Phe	Cys	Pro	Phe	Ser	Ile	Ile	Ser	215	220	225
Asn	Ala	Val	Ser	Asn	Ile	Ile	Cys	Ser	Leu	Cys	Phe	Gly	Gln	Arg	230	235	240
Phe	Asp	Tyr	Thr	Asn	Ser	Glu	Phe	Lys	Lys	Met	Leu	Gly	Phe	Met	245	250	255
Ser	Arg	Gly	Leu	Glu	Ile	Cys	Leu	Asn	Ser	Gln	Val	Leu	Leu	Val	260	265	270
Asn	Ile	Cys	Pro	Trp	Leu	Tyr	Tyr	Leu	Pro	Phe	Gly	Pro	Phe	Lys	275	280	285
Glu	Leu	Arg	Gln	Ile	Glu	Lys	Asp	Ile	Thr	Ser	Phe	Leu	Lys	Lys	290	295	300
Ile	Ile	Lys	Asp	His	Gln	Glu	Ser	Leu	Asp	Arg	Glu	Asn	Pro	Gln	305	310	315
Asp	Phe	Ile	Asp	Met	Tyr	Leu	Leu	His	Met	Glu	Glu	Glu	Arg	Lys	320	325	330
Asn	Asn	Ser	Asn	Ser	Ser	Phe	Asp	Glu	Glu	Tyr	Leu	Phe	Tyr	Ile	335	340	345
Ile	Gly	Asp	Leu	Phe	Ile	Ala	Gly	Thr	Asp	Thr	Thr	Thr	Asn	Ser	350	355	360
Leu	Leu	Trp	Cys	Leu	Leu	Tyr	Met	Ser	Leu	Asn	Pro	Asp	Val	Gln	365	370	375
Glu	Lys	Val	His	Glu	Glu	Ile	Glu	Arg	Val	Ile	Gly	Ala	Asn	Arg	380	385	390
Ala	Pro	Ser	Leu	Thr	Asp	Lys	Ala	Gln	Met	Pro	Tyr	Thr	Glu	Ala	395	400	405
Thr	Ile	Met	Glu	Val	Gln	Arg	Leu	Thr	Val	Val	Val	Pro	Leu	Ala	410	415	420
Ile	Pro	His	Met	Thr	Ser	Glu	Asn	Thr	Val	Leu	Gln	Gly	Tyr	Thr	425	430	435
Ile	Pro	Lys	Gly	Thr	Leu	Ile	Leu	Pro	Asn	Leu	Trp	Ser	Val	His	440	445	450
Arg	Asp	Pro	Ala	Ile	Trp	Glu	Lys	Pro	Glu	Asp	Phe	Tyr	Pro	Asn	455	460	465
Arg	Phe	Leu	Asp	Asp	Gln	Gly	Gln	Leu	Ile	Lys	Lys	Glu	Thr	Phe	470	475	480
Ile	Pro	Phe	Gly	Ile	Gly	Lys	Arg	Val	Cys	Met	Gly	Glu	Gln	Leu	485	490	495
Ala	Lys	Met	Glu	Leu	Phe	Leu	Met	Phe	Val	Ser	Leu	Met	Gln	Ser	500	505	510
Phe	Ala	Phe	Ala	Leu	Pro	Glu	Asp	Ser	Lys	Lys	Pro	Leu	Leu	Thr	515	520	525
Gly	Arg	Phe	Gly	Leu	Thr	Leu	Ala	Pro	His	Pro	Phe	Asn	Ile	Thr	530	535	540

Ile Ser Arg Arg

&lt;210&gt; 20

&lt;211&gt; 2603

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7248285CB1

&lt;400&gt; 20

```

atggcggtggg cccaccagc caccctcttt ctgttctgc tgetgctagg ccagccccc 60
ccagcagggc cacagtcact gggcaccact aagctccggc tgggtggccc agagagcaag 120
ccagaggagg gccgcctgga ggtgctgcac cagggccagt ggggcaccgt gtgtgatgac 180
aactttgcta tccaggaggc cacagtggct tggccccagc tgggcttoga agctgccttg 240
acctggggccc acagtgccaa gtacggccaa ggggaggggac ccactctggt ggacaatgtg 300
cgctgtgtgg gcacagagag ctecttgac cagtgcgggt ctaatggctg gggagtcagt 360
gactgcagtc actcagaaga cgtaggggtg atatgccacc cccggcgcca tcgtggctac 420
ctttctgaaa ctgtctccaa tgcccttggg ccccgaggcc agcggtgga ggaggtgagg 480
ctcaagccca tccttgccag tgccaagcag catagcccag tgaccgaggg agccgtggag 540
gtgaagtatg agggccactg gcggcaggtg tgtaccagg gctggaccat gaacaacagc 600
ggggtgtgtg gcgggatgct gggcttcccc agcgaggtgc ctgtcgacag ccactactac 660
aggaaagtct gggatctgaa gatgagggac cctaaagtct ggctgaagag cctgacgaat 720
aagaactcct tctggatcca ccaggtcacc tgcttgggga cagagcccca catggccaac 780
tgccaggtgc aggtggctcc agcccggggc aagctgcggc cagcctgccc aggtggcatg 840
catgctgtgt tcagctgtgt ggcagggcct cacttccgcc caccgaagac aaagccacaa 900
cgcaaagggt cctgggcaga ggagccgagg gtgcgcctgc gctccggggc ccaggtgggc 960
gagggccggg tggaagtgtc catgaaccgc cagtggggca cggctctgtg ccacaggtgg 1020
aaactcatct ctgccagtgt cgtgtgtcgt cagctgggct ttggctctgc tcgggaggcc 1080
ctctttgggg ccggctgtgg ccaagggtca gggcccatcc acctgagtga ggtgcgctgc 1140
aggggatatg agcggaacct cagcgactgc cctgccctgg aagggtccca gaatggttgc 1200
caacatgaga atgatgctgc tgtcaggtgc aatgtcccta acatgggctt tcagaatcag 1260
gtgcgcttgg ctggtgggag tatccctgag gaggggctat tggaggtgca ggtggagggtg 1320
aacgggggtc cagctgtggg gagegtgtgc agtgaaaact gggggctcac cgaagccatg 1380
ctggcctgcc gcagctcggc cctgggtttt gccatccatg cctacaagga aacotgggtc 1440
tggtcgggga cgccaagggc ccaggaggtg gtgatgagtg ggggtgcgctg ctcaggcaca 1500
gagctggccc tgcagcagtg ccagaggcac gggccgggtg actgctccca cgggtggcggg 1560
cgcttctctg ctggagtctc ctgcatggac agtgcaccag acctggtgat gaacgccag 1620
ctagtgcagg agacggccta cttggaggac cgcccgctca gccagctgta ttgtgccac 1680
gaggagaact gcctctccaa gtctgaggat cacatggact ggccctacgg ataccgcgc 1740
ctattgcgct tctccacaca gatctacaat ctgggcccga ctgactttcg tccaaagact 1800
ggaacgcata gctgggtttg gcaccagtgc cacaggcatt accacagcat tgaggctctc 1860
accactacg acctcctcac tctcaatggt tccaagggtg ctgaggggca caaggccagc 1920
ttctgtcttg aggacacaaa ctgccccaca ggactgcagc ggcgctacgc atgtgccaac 1980
tttgagaaac agggagtgtg ttaggctgct tgggacacct accggcatga cattgattgc 2040
cagtgggtgg atatcacaga tgtgggcccc ggggaattata tcttccaggt gattgtgaac 2100
ccccactatg aagtggcaga gtcagatttc tccaacaata tgetgcagtg ccgctgcaag 2160
tatgatgggc accgggtctg gctgcacaac tgccacacag ggaattcata ccagcccaat 2220
gcagaactct ccctggagca ggaacagcgt ctcagggaac acctcatctg aagetgtcac 2280
tgcacactcc tagctgctgc cgatacacca gatacctcag cttattggag ccattgccctt 2340
cacagagtcc caactcagag gaaaaggggc agtgcgaagg ggcaccaaga acctgctcag 2400
gaagcctttt gatggcaaga tcaccaatcc agatggtatt gctccctcag gatggctctg 2460
ggcctgcccc taagggcctg tggcctatgg aatatgtcct ccaggctttg ctgagctgag 2520
ctcctcttct gtaaggaaac ccagtcaccc ctgaatcttg ccacagagat ccgggattca 2580
ggagctctca gtttcttaag gag 2603

```

&lt;210&gt; 21

&lt;211&gt; 1745

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7472835CB1

&lt;400&gt; 21

```

gtaatttttt ttaaatggga tggatgacag tgacagagct ctaaattttt actaaccagc 60
tgagacacaa tggccaaaaa agcgattgct gtgattggag ctggaattag cggactgggg 120
gccatcaagt gctgcctgga tgaagatctg gagcccacct gctttgaaag aaatgatgat 180
attggacatc tctggaaatt tcaaaaaaat acttcagaga aaatgcctag tatctacaaa 240
tctgtgacca tcaatacttc caaggagatg atgtgcttca gtgacttccc tgtccctgat 300
cattttccca actacatgca caactccaaa ctcatggact acttcgggat gtatgccaca 360
cactttggcc tctgaatta cattcgtttt aagactgaag tgcaaagtgt gaggaagcac 420
ccagattttt ctatcaatgg acaatgggat gttgttggtg agactgaaga gaaacaagag 480
acttttgtct ttgatggggt cttagtttgc agtggaacac acacagatcc ctacttacca 540
cttcagtcct tcccaggat ggagaaaatt gaaggctggt atttccatag tcgggaatac 600
aaaagtcccg aggacttttc agggaaaaga atcatagtga tcggcattgg aaattctgga 660
gtggatattg cggtgagact cagtcgtgta gcaaacagg ttatattcct tagtactaga 720
cgtggatcat ggattttaca ccgtgtttgg gataatgggt atcccagga tagttcattt 780
ttcactcgtt tcaatgattt tctccagaaa atactaacta caccacaaat aaataaccag 840
ctagagaaaa taatgaactc aagatttaat catgcgcact gtggcctgca gcctcagcac 900
agagctttta gtcagcatcc aactgtcagt gatgacctgc caaatcacat aatttctgga 960
aaagtccaag taaagccag cgtgaaggag ttacacagaa cagatgccat ttttgaagac 1020
agcactgtag aggagaatat tgatgtgtgc atctttgcta caggatcacg tttttctttt 1080
tctttccttg atggtctgat caaggttact aacaatgaag tatctctgta taagcttatg 1140
ttccctcctg acctggagaa gccaaccttg gctgtcatcg gtcttatcca accactgggc 1200
atcatcttac ctatgcaga gctccaatct cgttgggcta cactgagttt caaaggctcg 1260
atcaaattac cctcagcgga gaacatgatg gcagatattg cccagaggaa aagggctatg 1320
gaaaaacggt atgtaaagac accccgccac acaatccaag tggatcacat tgagtacatg 1380
gatgagattg ccattgccagc aggggtgaaa cccaacctgc tcttctctct tctctcagat 1440
ccaaagctgg ccattggagg tttctttggc ccctgcaccc cataccagta ccacctccat 1500
gggcccagga aatgggatgg ggcccggaga gctaacctga cccagagaga gaggatcatc 1560
aagcccctga ggactgcagt tactagttag gacagccacc catctcaca gctctcttgg 1620
ataaagatgg ccccagtgag cctggcattt ctggctgctg gcttggcata ctttcgatat 1680
actccttacg gtaaatggaa ataaatgaaa gaacactgag ggggaaaagc atggaatagt 1740
ttcta 1745

```

&lt;210&gt; 22

&lt;211&gt; 1587

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7476203CB1

&lt;400&gt; 22

```

atgctctccc tgetcagtg gctggcttta ctggccatct ccttctctgt cctgaaactt 60
ggcaccttct gctgggacag gagctgtctt cctcctggcc cactccctt ccccatcctt 120
ggaaacctgt ggcagctatg ctttcagcag cctcaccttt cacttaaaaa ctttcagaag 180
aagataggaa atatctttat gaacttggga agcagtgtgg ttccactggc attgccattg 240
ttaccggtga ctttccatcc cttgaaccaa ggctgtttat gtaagccact tattactttc 300
cctaaacctt tccctaccag aaatccaggc atcatctgca gcagcgggca cactgtggcg 360
caaaagagac gcttctgcct ggtgatgatt cgagggtctg gcctaggcaa gctggcgctg 420
gaggtgcagc tgcagaaaga ggcagcagag ctggcagaag ccttccgcca ggagcagggt 480
aagagacctt tgcacctca ggtatccatt gtcaggteca cagtcagagt catcggggcc 540
cttgtgtttg gccaccaact cctcttagag gateccatct tccagggaact gactcaagcc 600
atcgactttg gcctggcatt tgtcagcact gtgtggcgcc agctgtatga cgtgtttccc 660
tgggcccctt gccacctccc aggacccccc caggagatat ttaggtacca agaggctctg 720
ctgagcttaa tccaccagga gatcaccagg cacaactca gggcacccga ggccccagg 780
gacttcacat gctgtacctt ggcccagatc tccaaggcca tggatgaccc tgtctccaca 840
ttcaaccagg agaacctggt ccagggtgtg atcgacctgt ttctgggagg caccgacacc 900
acagccacca ccctgtgctg ggcactcatc cacatgatcc agcacggagc tgtccaggag 960

```

```

acgggtgcagc tggagctgga cgaggtgctg ggtgctgccc egggtgtctg ctatgaagac 1020
cgcaagcgac tgccttacac catgctgtcc tccatgacgt gcagcgccctc agcagcgta 1080
tggccatggg tgcctgctgc cagtgtgtga cctccaccgc tgtgtgcagc tatcccgta 1140
gcaagggcac catcatctta cccaacctgg cctctgtgct ctatgacct gagtgtctgg 1200
agacccctcg acagttcaac cctggccact tctcggacaa ggatggaaac tttgtggcca 1260
atgaggcctt cctgccattc tctgcaggta ctagggtcta cccagcagac cagctggctc 1320
aaatggagct cttcctgatg tttgccaccc tctcaggac ctttcggttc caactgccag 1380
aagggagccc ggggctcaag ctggagtaca tctttggcgg cacttggcaa ccccgacccc 1440
aggagatctg cgcagtgcgc cgcctgagca gccccagccc tggctcctagg gaggatggcc 1500
tgtagccact gggggtctgg aggctgtcc cccatgaagt ccttcctcag tctcttttgg 1560
ttcctgcaaa gttagaaaag aggagga 1587

```

<210> 23  
 <211> 1038  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7478583CB1

```

<400> 23
atgaaagcag cagtttggtg tgggtcaaaa gatgtaagag ttgaagaacg tgaacctaaa 60
gaattacagg acaatgaagt taagggtaaa gtatcttggg ctggcatttg tggtagacat 120
ttacatgaat acttagaagg ccctatatat atttcaacag aaaagccaga tccattctta 180
ggtcaaaaag cgccagttac attgggtcat gaatttgcag gtgtagtaga agaaactggg 240
tcccaagtta caaaatttaa taaaggcgat cgagttgtag ttaaccctac agtttcaaaa 300
cgtgaaaaag aagaaaatat tgacctttat gatggttatt catctatagg cttagggtct 360
gtgggtggat ttgcagagtt tacaaatgcg ccggaagaaa atgtttataa actaccagat 420
aatgtttctg ataaagaagg tgcgcttgct gaaccaacag ccgttgcagt tcaagcaatt 480
aaagaagggt aagttctatt tgggtatact gtagctatct ttggtgcagg accaattgga 540
ttattaacag tcgtagcagc caaagcagct ggtgcaagta aaatatctgt tttcgattta 600
tcagaagaaa gactaagtaa agctaaagca ctaggcgcaa ctcatgctat aaactctggt 660
aagacagatc cagttgatgt tattaatgag tatacagaaa atggtgtaga tgtatctttt 720
gaagtggctg gtgtagcacc aacacttaaa tcttctatag atgttacaaa agcaagaggt 780
acagttgtta tcgtttctat ttttggctat cctatcgagt ggaatccaat gcaatttaact 840
aatacaggag taaaacttac ttctacaatt gcatacacac ctactacatt ccaacaaaca 900
attgacttaa tcaacgaagg taatttaaac gttaaagatg tagttactga tgaattgag 960
ttagaaaata tcgtagaatc aggatttgaa caacttgtaa atgataaatc tcaagcaaaa 1020
atattaatta aattataa 1038

```

<210> 24  
 <211> 1584  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7478585CB1

```

<400> 24
atgaaatcat taatgattgg cgctaattgt ggctcggtc aacatctcgt acgtaaacta 60
aatgcacgag atgttgattt tacggccggt gttagaaaag aagaacaagt tgaagcttta 120
aaagcagatg gtatcgatgc aacttacatt gatgttgcaa aacaatctat tgatgaatta 180
atagaattat ttaaaccgta tgaccaaatc cttttttctg tcggttctgg tggagcaca 240
ggtgacgac aaacaatcat agtagattta gacggttcag tgaagcaat taaagcaagt 300
gaacacgctg gtcgtcaaca ctttgttatg gtatcaacgt acgattcacg tcgagaagct 360
tttgatgctg caggcgactt gaaaccatac accattgcta aacattatgc ggatgactac 420
ttaagacatg caaatttaaa atataccatc gtacatccag gcgcattaac taacaatcat 480
gaaacgcaac aattcaatat gagtgtctaa tttgaaaatg tacaaaatcc gtctatcacg 540
agagaagatg tagcagaagt gcttgtttct gtgttaactg atgaaacatt acaagttgta 600
cttgcaaaac gaccacaaag tatccctcaa gacgatgtat ttagatttga aacaatagaa 660

```

```

actcgagaac caccatgcagg tgagggttcaa gtagaatcca tttatgtatc tgtagatcct 720
tacatgagag gcagaatgaa tgatacaaaa agttatgttc aacctttcca agtgaatgaa 780
ccattacaag gtcatattgt tggaaaagtc acacaatcga acgatgaacg tctatctgtt 840
ggcgattatg tcacaggcat attaccatgg aaaaagataa atacagtga tggagacgat 900
gtgacccctg tgcacatcaa agatgtacca ttacatttat atttgagtgt tttaggcatg 960
ccgggaatga cggcatatac aggattgctt caaattgggtc aaccacaatc tggcgagacg 1020
gttgtctgtt aacacttaaa aggtgcagta ggatctgtcg taggacaaat tgctaagatt 1080
aaaggcgcaa aagtgtcggc tattgtctgtt ggtaagcaga aaacaacata tttaacagat 1140
gaattagatg ttgatgcggc cattgactat aaacaagatg atttcgcaca gcaactcgaa 1200
gcggctgtac cagatgggat tgatgtgtat tttgaaaatg taggcggcgt aatttctgat 1260
gaagtgttta aacacttaaa tcgatttgca cgcgttcggg tatgtggtgc aatttcagca 1320
tataataatg aaaaagacga tattggacca cgtatccaag gaacgttgat taaaaatcaa 1380
gcattgatgc aagggtttgt agtagcaca ttcgctgatc attttaaga agcaagcgaa 1440
caactcgac aatgggtgtc tgaaggtaaa attaaatttg aagtgcgat agatgaaggt 1500
tttgacaatt taactctgc attcagaag ttatttacag gtgataattt cggtaacaa 1560
gttgtcaaaa tcaagaaga atag 1584

```

&lt;210&gt; 25

&lt;211&gt; 1500

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7479904CB1

&lt;400&gt; 25

```

atggatgaga aatccaacaa gctgctgcta gctttggtga tgctcttccct atttgcctg 60
atcgtctctc aatacgtgtg ccccggcaca gaatgccagc tcctccgcct gcaggcgttc 120
agctccccgg tgcgggaccc gtaccgctcg gaggatgaga gctccgccag gttcgtgccc 180
cgctacaatt tcacccggcg cgacctcctg cgcaaggtag acttcgacat caaggggcat 240
gaectgatcg tgttcttga catccagaag accgggggca ccactttcgg ccgccacttg 300
gtgcgttaaca tccagctgga gcagccgtgc gagtgccgcy tgggtcagaa gaaatgcact 360
tgccaccggc cgggtaagcg ggaaacctgg ctcttctcca gggtctccac gggctggagc 420
tgccgggtgc acgcccactg gaccgagctc accagctgtg tgccctccgt ggtggacggc 480
aagcgcgacg ccaggctgag accgtccagg tggaggattt ttcagattct agatgcagca 540
agtaaggata aacgggggtt tccaacact aacgcaggcy ccaactctcc gtcctccaca 600
aagaccggga acacatctaa gagtgggaag aacttccact acatcccat cctccgagac 660
ccagtgtccc ggtacttgag tgagtggagg catgtccaga gaggggcaac atggaaagca 720
tccctgcctg tctgcgatgg aaggcctcca acctccgaag agctgcccag ctgctacact 780
ggcgatgact ggtctggctg cccctcaaaa gagtttatgg actgtcccta caatctagcc 840
aacaaccggc aggtgcgcac gctctccgac ctgacctgg taggctgcta caacctctct 900
gtcatgcctg aaaagcaaa aaacaaggtc cttctgaaa gtgccaagtc aaatctgaag 960
cacatggcgt tcttcggcct cactgagttt cagcggaaga cccaatatct gtttgagaaa 1020
accttcaaca tgaactttat ttcgccattt acccagtata ataccactag ggccctctagt 1080
gtagagatca atgaggaaat tcaaaagcgt attgagggac tgaattttct ggatatggag 1140
ttgtacagct atgccaaa cctttttttg cagaggatc agtttatgag gcagaaagag 1200
catcaggagg ccaggcgaaa gcgtcaggaa caacgcaaat ttctgaaggg aaggctcctt 1260
cagaccatt tccaggacca gggtcagggc cagagccaga atccgaatca gaatcagagt 1320
cagaacccaa atccgaatgc caatcagaac ctgactcaga atctgatgca gaatctgact 1380
cagagtttga gccagaagga gaaccgggaa agcccgaagc agaactcagg caaggagcag 1440
aatgataaca ccagcaatgg caccacgcac tacataggca gtgtagagaa atggcggtta 1500

```

&lt;210&gt; 26

&lt;211&gt; 669

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7480367CB1

&lt;400&gt; 26

```

atggcagaga agcccaagct ccactactcc aatgcacggg gcagtatgga gtccattcgg 60
tggctcctgg ctgcagctgg agtagagttg gaagagaaat ttctagaatc tgcagaagat 120
ttggacaagt taagaaatga tgggagtttg ctgttccagc aagtaccaat ggttgagatt 180
gacgggatga agctggtgca gaccagagcc attcttaact acattgccag caaatacaac 240
ctttatggga aagacatgaa ggagagagcc ctgattgata tgtacacaga aggtatagta 300
gatttgactg aaatgatcct tcttctgctc atatgtcaac cagaggaaag agatgccaag 360
actgccttgg tcaaagagaa aataaaaaat cgtacttccc ctgcctttga aaaagtatta 420
aagagccaca gacaagacta ccttggttgc aacaagctga gctgggctga cattcacctg 480
gtggaacttt tctactacgt ggaagagctt gactcgagtc ttatctccag cttccctctg 540
ctgaaggccc tgaaaaccag aatcagcaac ctgcccacgg tgaagaagtt tctgcagcct 600
ggcagccaga gaaagcctcc catggatgag aaatctttag aagaagcaag gaagattttc 660
aggttttaa
669

```

&lt;210&gt; 27

&lt;211&gt; 3551

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 8069390CB1

&lt;400&gt; 27

```

ggcgcgaggc agctccggcg gcgagacggg ggcgcgcccg cgcgggtctg gcgggacggg 60
tttgaagac tttgccggcc tgcagattgg ccttaagaga aggacggagc cacatactgc 120
tgacggccca gaactggcag agagaaggtt gccatggctg ctggtgacag tttctacctc 180
ttgtacaggg aaatcgccag gtcttgcaat tgctatatgg aagctctagc tttggttggg 240
gcctggtata cggccagaaa aagcatcaact gtcatctgtg acttttacag cctgatcagg 300
ctgcatttta tccccgcctt ggggagcaga gcagacttga tcaagcagta tgggaagatgg 360
gcggtgtgca gcggtgcaac agatgggatt ggaaaagcct acgctgaaga gttagcaagc 420
cgaggtctca atataatcct gattagtcgg aacgaggaga agttgcaggt tgttgctaaa 480
gacatagccg acacgtacaa agtggaaact gatattatag ttgcggactt cagcagcggg 540
cgtgagatct accttccaat tcgagaagcc ctgaaggaca aagacgtttg catcttggtg 600
aataacgtgg gtgtgtttta tccctacccg cagtatttca ctacagctgc cgaggacaag 660
tcctgggaca atataatgt gaacattgct gcgctagtt tgatgggtcca tgttggttta 720
ccgggaatgg tggagagaaa gaaagggtgcc atcgtcacga tctcttctgg ctccctgctgc 780
aaaccctctc ctacagctggc tgcattttct gcttctaagg cttattttaga ccacttcagc 840
agagccttgc aatatgaata tgcctctaaa ggaatctttg tacagagtct aatcccttcc 900
tatgtagcca ccagcatgac agcaccagc aactttctgc acaggtgctc gtggttgggt 960
ccttcgcca aagctcatgc acatcatgct gtttctactc ttgggatttc caaaaggacc 1020
acaggatatt ggtcccatcc tattcagttt ctttttgca agtatatgcc tgaatggctc 1080
tgggtgtggg gagcaaatat tctcaaccgt tcaactacgt aggaagcctt atcctgcaca 1140
gcctgagctt ggtggccac ttgagaagtt ttgccaactc ctgggaacct cgatattctg 1200
acatttggaa aaacacattt aatttatctc ctgtgtttca ttgctgatta ttcagcatac 1260
tgttgattcg tcatttgcaa aacacacata ataccgtcag agtgcgtgta aaaaccttaa 1320
gggtgtgtgg atggcacagg atcaataatg cctgaggctg attgacgaca tctacatttc 1380
agtgcctttt ccctaagctg tttgaaagtt acgcttttct gttgttctag agccacagca 1440
gtctaataat gaaatataat atgatttgto aggtcttata atttcagatg ttgtttttta 1500
agggaaattg accatttcac tagaggagtt gtgctggttt ttaaatgtgc atcaagaaag 1560
actactgaaa agtattattt tgtaactaag attgctggta ctattaggaa aaatctgtgt 1620
gtattgtata gctctagctg tttgactatc tgtaatgaaa atgctgcact tcaactggta 1680
tttcattaga gaaccgtgtg tgtgcgtgtg tgtgggtgct ttgagcaact ttatttatgg 1740
ttaccatatt tttaaaaga ttttttgcga ggggtgactta acatggactc ttatagggtg 1800
ttaaacaat ctagattatt ctttttcatc ctaaaataag ctaccaaatt tcatgctgtt 1860
ggtttgccat gaatgatatt acttctaca ttatatttgt gttttttcaa atctgctatg 1920
gaatgaactt attcctagat ttggatatgt aagagaaacc tgcagtcate ttttgattta 1980
taaggaattt cttgtggata aatagtgatt tctcagctc tgacccattt tataactgaa 2040
atttagccct ttagagcttg ttatctctgg ttttctacg tttttctatg taatattatt 2100
ccattccagt agcattattg atagaaatag taagtattta tggaaatagta aaatatggac 2160
aaattacgtg tgtgacatat ctgtcaaaat aagttagaag ctattctctg gtttgtgtaa 2220
tgaatttatg tattgtagtg aataccttta ctggtgtgaa gataattatg cacaaccctc 2280

```

```

cacaatacgc gttaacattg aaacctgtga aatgtcctta ggttgggtca tataaagcca 2340
accatttttg aggaccatgt acctagtgtt ttgaaaactg taagtcacta tatgaatatg 2400
acaatatgtg cacattttaaa attcagagct cggcattgtg atactgatgc agaagctagt 2460
agattgggta aaagtctgga cttctgtggc atttttttcg tgacgtgata atctatcata 2520
agcagaccta agcacagttt tatgaacaca attttgcca tgacattgcc tacaggattt 2580
ccagatgtga cttgcaactca gaagatcagt ggtcaacttc agaagttctt ccacgcttag 2640
atcatgtctt cagaacttag atgtgaaaat ctacacactg ggagatgctg tgagcccca 2700
ggttttgatg gagtttgctt ggaatcctct tgacttcctg ccacattgac gtgaactttg 2760
atgtataata agcagcagca acttcctgtg aaaatatggt caggtagtta tatgtaaggt 2820
tacgtggtcc agtaatgtct tagattgata aattaggtat ggaatccatc agtggtacgt 2880
gatgagaata ggtgaacaca cttgtcagc gatgatgtaa acttctctcc ttggcaggac 2940
atgggcaaac atgctgattg gtgcaaatgt ggtgccgagc tgtccatagc tgcagtga 3000
gatgaagagc aagacctctt ctaggttttc tagctttcat taaatgtatt tttttcccca 3060
gagctaattt gaaagttgat tggaccactg tggatggggt ctcattaaaga atgtgggaaa 3120
taggggcccga gtgcggtggc tcacacctgt aatcccagca gtttgaagg ccaggggcagg 3180
tggatcgctt gatccaggga ggtcgagacc agcctgggga acacatcctg tctctacaaa 3240
aaatacaaaa attagccagg cagggtggtg catgcctgta gtcccagcta cttgggaggc 3300
tgaggcagga gaattttttg agcccaggat gcagagggtg aagtgaacca agatcgtgcc 3360
actgcactcc agccttgaga cagagcgaga cctgtctca aaaaaaaaaa aagaacgtgg 3420
gaaatatgaa ctttgaaag ttaatctgtg aattgaaagt ttaacaataa aagtagttgt 3480
ttgtttcctt tgaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaatggcgg 3540
tcgcaagctt a 3551

```

&lt;210&gt; 28

&lt;211&gt; 2178

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7473869CB1

&lt;400&gt; 28

```

tcgggtgggg agtagagtag gtaagcgtgt ttctttattg ggggagggtg tgtgtaaccg 60
agagggtttc actgccaatt atggaagtag cctctccttg atattgttga ggtgggattg 120
tgcaacaaca ttgtactgtt gtgggttgta atttcgccta gatggcataa gtgagaactg 180
aagacttcag gtaggacag aggcacgggt cttgttagtt atgctcttgg gatcagtgtc 240
gcctgctgag ctgaaaagga aatggattca atctcttcaa cctttaagggt gatagatagt 300
ttgagcagac tggagaatgg acacactatg aagctgtggc tagaaagggg ctgggtcatgt 360
cccactctct ggcagagatt actggggatg tccggacaga tgcctgcatt ggtgggtgagg 420
gccacatctg cacacgagcc agtggctgct tgcagttcac tgcctgtgat ccagagtgtg 480
ttcaaagggtg actctcctgc tcttctggac tcttctctca ggcaagaaag gctgcaggct 540
gcctgctatg tgatgcctga gcacaaagcc aaggaactga actaagtctt tctgttaagt 600
cctgagtttg tcattggcag gtttacttgt ggccagctct cctgcccctt ggggggttccg 660
tcttctcact ggcgacctg gattgaaacg atctcccgc ggccgcgcgc gctacctggg 720
gcccgcagggt gctgcagga gtccctgggc cagctggcct cgatgtacgt cagcacgcgg 780
gaacggtaca agtggctgag cttcagcgag gactgtctgt acctgaacgt gtacgcgcgg 840
gcgcgcgcgc cgggggatcc ccagctgcca gtgatggtct ggttcccggg aggcgccttc 900
atcgtgggag ctgctctctt gtacgagggc tctgacttgg ccgcccgcga gaaagtgggt 960
ctgggtgtttc tgcagcacag gctcgccatc ttccgcttcc tgagcacgga cgacagccac 1020
gocgcgcggga actgggggct gctggaccag atggcggtc tgcgtgggt gcaggagaac 1080
atcgagcct tcggggggaga cccaggaaat gtgacctgt tccgccaagt ggcggggggc 1140
atgagcatct caggactgat gatgtcacc ctagcctcgg gtctcttcca tcgggccatt 1200
tcccagagtg gcaccgctt attcagactt tcatcacta gtaaccact gaaagtggcc 1260
aagaagggtt cccacctggc tggatgcaac cacaacagca cacagatcct ggtaaactgc 1320
ctgagggcac tatcagggac caaggtgatg cgtgtgtcca acaagatgag attcctcaa 1380
ctgaacttcc agagagaccc ggaagagatt atctggtcca tgagccctgt ggtggatgg 1440
gtggtgatcc caagtgacct tttggtgctc ctgaccagg ggaaggtttc atctgtgcc 1500
tacctctag gtgtcaacaa cctggaattc aattggctct tgccttatat catgaagttc 1560
ccgctaaacc ggcaggcgat gagaagga accatcacta agatgctctg gagtaccgcg 1620
accctgttgg tgagggaccc agctggcagg ggtgctcagt tccgacaggg ttgaccccc 1680
tgttttttt aacctagtag ctgctctttg caaaggggct ccagccagg gtaaggatct 1740

```

```

ttcttgagg gtctggggtt tgetgtggga tcagatgact gcttacaggt aagggtgctca 1800
gggtcacagg ggcagttatg cagcaaaatc aggggttaca atcagcagag acagaaactt 1860
tcccaggaag ctccctttct cccctccca ggcacaaaac tcctgggggg ctgagcatgg 1920
atccaagtca ctggtgggccc cacctctggc ccagctggca cccaggcctc aggttaagtgt 1980
ggcctcctca ctcagaatat caccaaggag caggtaccac ttgcgggtgga ggagtacctg 2040
gacaatgtca atgagcatga ctggaagatg ctacgaaacc gtatgatgga catagttcaa 2100
gatgccactt tegtgtatgc cacactgcag actgctcact accaccgaga tgccggcctt 2160
cctgtctacc tgtatgaa 2178

```

```

<210> 29
<211> 2081
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7478588CB1

```

```

<400> 29
gctgtgggct ggtcagaagc tgggtacaat tccccccgcc ccagtacttg ctggcaggga 60
ttaagagcag ataaaaagtgt gtcacacac tgtagacacg gctaccatgc catccacagt 120
gttgccatcc acagtgttgc catcactcct gccacagca ggagctggct ggagcatgag 180
gtggattctg tgetggagcc tcacctctg cctgatggcg cagacggcct tgggtgcctt 240
gcacaccaag aggcctcaag tggtcaccaa atatggaacc ctgcaaggaa aacagatgca 300
tgtggggaag acacccatcc aagtcttttt aggagtcccc ttctccagac ctctctagg 360
tatctcagg tttgcacctc cagaaccccc ggagcccttg aaaggaatca gagatgctac 420
cacctaccgg cctgggtgcc tgcaggagtc ctggggccag ctggcctcga tgtacgtcag 480
cacgcgggaa cgggtacaagt ggctgcgctt cagcgaggac tgtctgtacc tgaacgtgta 540
cgcgcgggag cgcgcggccc gggatcccca gctgccagtg atgggtctggt tcccgggagg 600
cgccttcacg gtgggcgctg cttctctgta cgagggtctt gacttggcgg cccgcgagaa 660
agtgtgtctg gtgtttctgc agcacaggct cggcatcttc ggcttcttga gcacggacga 720
cagccacgag cgccgggaact ggggctgct ggaccagatg gggctctgca gctgggtgca 780
ggagaacatc aggcctctcg ggggagaccc aggaatgtg accctgttct gccagtcggc 840
ggggggccatg agcatctcag gactgatgat gtcaccccta gctcgggtc tcttccatcg 900
ggccatttcc cagagtggca ccgcgttatt cagacttttc atcactagta acccactgaa 960
agtggccaag aaggttgccc acctggctgg atgcaaccac aacagcacac agatcctggt 1020
aaactgcctg agggcactat cagggaccaa ggtgatgcgt gtgtccaaca agatgagatt 1080
cctccaactg aacttcaga gagaccgga agagattatc tgggtccatga gccctgtggt 1140
ggatggtgtg gtgatcccaag atgacctttt ggtgtcctg acccagggga aggtttcatc 1200
tgtgccttac cttctagggt tcaacaacct ggaattcaat tggctcttgc cttatatcat 1260
gaagtctccg ctaaaaccggc aggcgatgag aaaggaaaac atcactaaga tgctctggag 1320
taccgcacc ctgttgaaata tcaccaagga gcaggtagca cttgtggtgg aggagtacct 1380
ggacaatgtc aatgagcatg actggaagat gctacgaaac cgtatgatgg acatagttca 1440
agatgccact ttctgtatg ccacactgca gactgtcac taccaccgag atgcccggct 1500
ccctgtctac ctgtatgaat ttgagcaca cgctcgtgga ataactgtca aaccocgcac 1560
tgatggggca gacctgggg atgagatgta cttctctttt gggggccctc tcgccacagg 1620
cctttccatg ggtaaggaga aggcacttag cctccagatg atgaaatact gggccaactt 1680
tgcccgcaca ggaaacccca atgatgggaa tctgccctgc tggccacgct acaacaagga 1740
tgaaaagtac ctgcagctgg attttaccac aagagtgggc atgaagctca aggagaagaa 1800
gatggccttt ttgatgagtc tgtaccagtc tcaaaagacct gagaagcaga ggcaattcta 1860
aggttggtcta tgcaggaagg agccaaagag gggtttgccc ccaccatcca ggccctgggg 1920
agactagcca tggacatacc tggggacaag agttctaccc accccagttt agaactgcag 1980
gagctccctg ctgcctccag gccaaagcta gagcttttgc ctgttgtgtg ggacctgcac 2040
tgccctttcc agcctgacat cccatgatgc ccctctactt c 2081

```

```

<210> 30
<211> 2642
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature

```

&lt;223&gt; Incyte ID No: 55046125CB1

&lt;400&gt; 30

```

gagccggccc gggctcggctcg ccgcctcagtt tcgcgcggggc cctccctaggg gtgtgtctca 60
eggattccac acccggccgc tcctggacaa gccccgaaag gcgtcttctt ccttggcggg 120
agccgcgtgc gcccgcgttt tcgcgtgct gtccccgggc cgcgcaggc ggatgcacgt 180
cctcaggcga cgctgggacc tgggtccct ctgcggggc ctgctcactc ggggcctggc 240
cgccctgggc cactcgctga agcacgtgct cgggtcgatc ttctccaaga ttttcggccc 300
catggccagc gtogggaaca tggatgagaa atccaacaag ctgctgctag ctttgggtgat 360
gctcttctta tttgcgtga tcgtctccca atacgtgtgc cccggcacag aatgccagct 420
cctccgcctg caggcggttca gctccccggt gccggaccgc taccgctcgg aggatgagag 480
ctccgccagg ttcgtgcccc gctacaattt caccgcgggc gacctcctgc gcaaggtaga 540
cttcgacatc aaggggcgatg acctgatcgt gtctctgcac atccagaaga cggggggcac 600
cactttcgcc cgccacttgg tgcgtaacat ccagctggag cagccgtgcg agtgccgcgt 660
gggtcagaag aaatgcactt gccaccggcc gggtaagcgg gaaacctggc tcttctccag 720
gttctccacg ggtctggagct gcgggttgca cgcgcactgg accgagctca ccagctgtgt 780
gccctccgtg gtggacggca agcgcgacgc caggctgaga ccgtccaggc acttccacta 840
catcaccatc ctccgagacc cagtgtcccg gtacttgagt gagtggaggc atgtccagag 900
aggggcaaca tggaaagcat cctcgatgt ctgcgatgga aggcctccaa cctccgaaga 960
gctgccagc tgetacactg gcgatgactg gtctggctgc cccctcaaag agtttatgga 1020
ctgtccctac aatctagcca acaaccgcca ggtgcgcatg ctctccgacc tgacctggt 1080
aggctgctac aacctctctg tcactgcctga aaagcaaaag aacaagggtcc ttctggaaaag 1140
tgccaagtca aatctgaagc acatggcggt ctccggcctc actgagtttc agcgggaagac 1200
ccaatatctg tttgagaaaa ccttcaacat gaactttatt tcgccattta cccagtataa 1260
taccactagg gcctctagtg tagagatcaa tgaggaaatt caaaagcgta ttgagggact 1320
gaattttctg gatatggagt tgtacagcta tgccaagac ctttttttgc agagggtatca 1380
gtttatgagg cagaaagagc atcaggaggc caggcgaaag cgtcaggaa aacgcaaatt 1440
tctgaaggga aggtcctctc agacccattt ccagagccag ggtcaggggc agagccagaa 1500
tccgaatcag aatcagagtc agaaccctaa tccgaatgcc aatcagaacc tgactcagaa 1560
tctgatgcag aatctgactc agagtgttag ccagaaggag aaccgggaaa gcccgaaagc 1620
gaactcaggc aaggagcaga atgataacac cagcaatggc accaaccgact acataggcag 1680
tgtagagaaa tggcgttaaa tggctcaaaa aggcctgtac atacttctcc caaagcgcca 1740
ctgaaaagat ggcatagett aaaagatgaa agtgtccaaa cacatcctgc ttccttcatt 1800
ggggaggttt taaaaaaaag tttagatgtt gcctttacag ttgcctttca attcagtgtt 1860
atactgtgtg taggtaaaac aaatctcaat atggaattaa attgtctttt tgggggttga 1920
ctaaatatga aatccgaaag ccaaaccaga ctaccagaa attgctgttt agatatttta 1980
agaagttctt aaattagtta tggagacaaa gtgaaaacat aaaatgtgac catttaactt 2040
atggctaaga aatggacttt aaattattca tgatacactg ttaaaacca atcttggat 2100
caaatatttt ttccaggggt gagaataagt ataaacataa agcaactaaa atgaaacata 2160
aaacctttta ttttctctct attttaacaa ggaatctatt taaatagaat aacaactgat 2220
gggtaactct accgagctgt agaaaataaa aaattcctct ccaaacatgg gtagttttat 2280
gtcaaaatat tggcttttca agaacaggac tcatactctg atatttaaga gatgtttaaa 2340
attttaaaact ttttctacct tctactgttt aaagggttta cacagggtgt atctcacatt 2400
aaacaaaaca ccttttttcc aaaatgaaat accaatgtaa agatctaatt tccaggcgct 2460
ttcagggcac tgtaatttca acaatactgg aatcattttg gcgtgcttc tcattcattt 2520
taaggcttct ctgaattgtg ctcatcccaa attaacccat gtatagaatc tttcttcate 2580
atctaaatgg tgtgttgctg aagtatttgt ggtatataat cctggattaa agtcaggact 2640
tt 2642

```

&lt;210&gt; 31

&lt;211&gt; 2080

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 3538709CB1

&lt;400&gt; 31

```

tgttttcggcg gccgcgggat gccctgcgc tgaccgccag gggcagggtgc ccgcccgcgt 60
agacgcaccc ggcctgacct cgcgccacca tgtaaacagc gccagcaggc ggacgctggc 120
ttctccgctt gggacccttc cgcgccgacc cgggccccgc ggcctcgat gaggacacac 180

```

```

catgctgacc ggggtgaccg acggtatctt ctgttgcttg ctgggcacgc ccccaacgc 240
cgtggggcca ctggagagcg tccagtcag cgatggctac acctttgtag aggtcaagcc 300
cgcccgctg ctgggggtga agcatgcagg acccgccccca gccgtgccc caccctccacc 360
atcatccgca tctcggatg cagcccagg ggacctctcc ggcttggtcc gctgtcagcg 420
ccggatcacc gtgtaccgca atggggcggt gctgggtgaa aacctgggcc gagccctcg 480
agccgacctc ctacacgggc agaattggctc tggggagccg ccggccgcc tggaggtgga 540
gctggcagat ccggcgggca gcgatggccg cttggccccc ggacgcgcag gcagcggcag 600
cggcagtgcc agtgggtggc ggccggcgcc agccaggcgc cccaagagga ccatccatat 660
tgactgtgag aagcgcatca ctactgtcaa aggcgccccag gccgacgtgg tgctcttttt 720
catecatggt gtccggcggt ccttgcccat ctggaaggag cagctggact tctttgtgag 780
cctaggctat gaggtggtg ctctgacct ggccggccac ggggccagct ctgcgcccc 840
gggtggccgca gctacacct tctatgcgct ggctgaggac atgcgagcaa tcttcaagcg 900
ctatgccaa aagcgaaatg tctcattgg ccattcctac ggtgtctctt totgcacatt 960
cctggcacat gagtaccag acctagtcca caaggtgatc atgatcaatg gggggggccc 1020
tacggcgctg gagccagct tctgtcaat cttcaacatg cccacctgag tctgtcactg 1080
cttgcgccc tgcctggcct ggagcttctt caaggccggc ttcgcccgc aaggagccaa 1140
ggagaagcag ctgttaaagg agggcaacgc tttcaacgtg tcatccttc tactccggcg 1200
catgatgagc ggccagtact ggcccaggcg cgacgaggtc taccacgcgc agctcaccgt 1260
gccgctcctg ctgttccacc gcatgcacga taagtttgtg ccggtggagg aagaccagcg 1320
catggccgag atcctgctcc tggcattcct gaagctcacc gacgagggca gccacatggt 1380
gatgctggaa tgccctgaga cgggtcaacac gctgtccac gaattcctgc tctgggagcc 1440
cgagccctcg cccaaggctc taccggagcc actgcggcg cctccagaag acaagaagta 1500
gccgctggcg ccggcgggga tccgttgggt agcacagccg cagcaggagg agggccgagc 1560
ctgcgccagg tctgcagcgc agaccacctg ggccggccgt tccctccggt gggcggggcc 1620
aggtcaggga gacgccccca ggctgcctgg gcggggcggt gcatccgagg gagccagcg 1680
gacattccgc tctccgcttc cgtcccgcg ggcccatcgg cgttttgggg ccgcagccgg 1740
gacctcacc gaagatgacc ttgtacagaa gctctcctc acctccccc caacgccacg 1800
gccaaggcag gccccccacc ccgctgtctt cagtgtcagc cgtgcttgat cctgggaccc 1860
acgagcccca cagggacctc cgaggcccca tccgcttacc cgagacctt cctaccccc 1920
attctcggc gctgggagct atttttgccc aagggggggg gatggggggg ctggcgccac 1980
cgaacctgca catctcaact tgtaactcaa taaacagaag tgacaatcgg aaaaaaaaa 2040
aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 2080

```

&lt;210&gt; 32

&lt;211&gt; 2219

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 71563101CB1

&lt;400&gt; 32

```

ggcctccaat ctgcttccat gggggttggc tttctgagtg ggagaaatga ctctaacttg 60
gagacatttg ctgagacct tgtgectggc cacttccgct cccaggatcc ttgagatgca 120
tcttttcttg agcctaggt cttcccgac atcagtaacc aagctcagtc ttcatacaaa 180
gccagaaatg cctccatgtg acttcatgcc tgaagatac cagtcccttg gctacaaccg 240
tgtcctggaa atccacaagg aacatcttcc tctgtgtgtg acggcatatt tccagaaacc 300
cctgctgctc caccaggggc acatggagtg gctctttgat gctgaaggaa acagatacct 360
ggatttcttt tccgggattg ttactgtcag tgttggccac tgccacccaa aggtgaatgc 420
agtggcacia aagcagctcg gccgctgtg gcatacaagc accgtcttct tccacctcc 480
aatgcatgaa tatgcagaga agcttgccgc acttcttctt gagcctctta aggtcatttt 540
cttggtgaac agtggctcag aagccaatga gctggccatg ctgatggcca gggcgccact 600
aaacaacata gacatcattt ctttcagagg agcctaccat ggatgcagtc cttacacact 660
tggcttgaca aacgtaggga tctacaagat ggaactcctt ggtgggacag gttgccaacc 720
aacaatgtgt ccagatgttt ttctgtggcc ttggggagga agccactgtc gagattctcc 780
agtgaacaca atcaggaagt gcagctgtgc accagactgc tgccaagcta aagatcagta 840
tattgagcaa ttcaagata cgctgagcac atctgtggcc aagtcaattg ctggattttt 900
cgcagaacct attcaagggt tgaatggagt tgtccagtac ccaaaggggt ttctaaagga 960
agcctttgag ctgtgtcgag caaggggagg cgtgtgcatt gcagatgaag tgccagacgg 1020
atttggaagg ttgggctctc acttctgggg cttccaaacc cagcatgtcc tgccctgacat 1080
tgtcaccatg gctaaaggga ttgggaatgg ctttcccatg gcagcagtc taaccactcc 1140

```

```

agagattgcc aatcttttgg cgaaatgcct gcagcacttc aacacctttg gagggaaacc 1200
catggcctgt gccattggat ctgctgtgct tgaggtgatt aaagaagaaa atctacagga 1260
aaacagtcaa gaagtttgga cctacatgtt actaaagttt gctaagctgc gggatgaatt 1320
tgaaattgtt ggagacgtcc gaggcaaagg cctcatgata ggcatagaaa tgggtgcagga 1380
taagataagc tgtcggcctc ttccccgtga agaagtaaat cagatccatg aggactgcaa 1440
gcacatggga ctccctcggtg gcagaggcag cattttttct cagacatttc gcattgcgcc 1500
ctcaatgtgc atcactaaac cagaagtiga ttttgcagta gaagtatttc gttctgcctt 1560
aaccacaacac atggaaagaa gagctaagta acattgtcag aaataaataa aaccacaagt 1620
ctcaagaatt tgcacagtat gttcaagggt gaatttgaag aatttcagaa ccaactggat 1680
ccagagaaag cctgcagctc tccacaggag ctgtaaaagt catggttgac tgcctaccaa 1740
ccatatttgt tagcagagcc cctcttatct tgagaactcc attcttcagg gaaaggatct 1800
ccctagctca gagaataaat cctaattagt ttatgttagg tatggtaatt tgattccctt 1860
ttgcagtgat tggtttatgc atgaatatgt gatgtatttt tgtccagtga atcttgaaga 1920
aaaatctttt ggtggagggt ccttcaggga aagttttctt caccctcact cttcagttca 1980
agaagagatg tcttcttgtt gcgctgagaa caccatatgt tcatgacgag attcctggca 2040
ccatgtcagc cggtttgtag tcatgaggac aacccttttt ggtgagggtg gaagatggat 2100
ggaagccaag tgettagtga tgtcaaagaa gcactcactt aagcattcct ggagccacc 2160
tacctcaggg cctcttgata tttgaggtaa taaaattcat tgttctgtat aaaaaaaaa 2219

```

&lt;210&gt; 33

&lt;211&gt; 681

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7472027CB1

&lt;400&gt; 33

```

atgagactat gtgaaaagac cgaactacag ttgattgggt tacctgaaag tgacagggag 60
aatggaacca agttgaaaaa cacatttcag gatattatcc aggagaactt cccaaccta 120
gcaagacagg ccaacattca aattcagatg gctgggtgat ctatctggat tgaagggatt 180
ccttttccca gcaataatth tacagacctg agacgtttgc aagatgaaat tgtgttgcgg 240
gatgaagatg tcattacact ttcttaccba aagtcaggaa gcttttggat agtggaaatc 300
atcagtctga tccactccaa gggagatcct agttgggtcc aatctgttgt tccctgggat 360
cgttcaccat ggatagaagt taaacgtaag aaagcagggt tagagagtca gaagggcca 420
cactctaca cctccacct tcccattcag ctcttcccca agtcattctt gaattccaag 480
gccaagtgtg tttatcctca tgttctcatg cttgtggttc tcatcctagg acataagagc 540
cagtggagta ttgtatataa gatattctgaa aatgcagaag caacttcaaa actgggtaat 600
gggcaaagat tggaaagatt tggagggctc agaagaagac aggaagatga aagaagtttg 660
gaatttctta gagactgtta a

```

&lt;210&gt; 34

&lt;211&gt; 399

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7480358CB1

&lt;400&gt; 34

```

agccagagcc agagtgtggc ccctggagtc ataagtatga aggtcatgta catgttgaag 60
ggccagagcc cgggtgcagg caccatccac tttgagcaga aggaataatga accatttatg 120
gtgtcagaat gcattacagg attgactgaa cgccagcaca gattccatgt tcatcagttt 180
ggagataata caccaggctg taccagggca gttccttact ttaatccttt aaccaaaaac 240
cacagtgggc caaggatcaa gagaggcagg ttggagacct gggtaatgtg gccgtggca 300
aagatgtgtc gccacatgtc tgttgaagat tctctgggtc cactctcagg acactattcc 360
atcactgccc acacaatggt gtccatgaca accagatga

```

&lt;210&gt; 35

&lt;211&gt; 2302

<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 1618256CB1

<400> 35  
taactaatcg gaaggcgetg tgaatttcac gttttctcgcg tagctgtaac ggcgggtata 60  
gttcttactc aagtcgtgcc ctgccagaaa ggggaagatg cagcgaggga gagacctaga 120  
accgcgatca aagctgcagc agctgctggg gctggcggca caaagggagg agggaggagc 180  
ctgggcgctg agaaagttct tggggaaggt tgagctgagc caagagtcgg gcggtggctg 240  
ggatgggagg gaggggcccg cccgatttcc cttgctgctc cccttggtgc ctgacgctga 300  
cagaggcaaa aatctgctaa ctcagggggc agactcaacc aagactgtga gcaggcctgg 360  
ggaatgaccc cccgactctcc aaccagtgcc tlccgcagct gcacggctgt ctccagctgt 420  
ctctgcccct ctctctggcc ctggctccat ctctctgtca cctcaccctt cctctgtcca 480  
catgggcccct ctctctccct ccaggacgct gcggctctgg ggacctegga gcctgggggt 540  
ggctctggga gtcttcatga ccattggctt tgcactccag ctcttgggag ggccttcca 600  
gaggaggcta gctgggctc agctccgaca gccctcgccc ccactccctac gaccagccct 660  
tccgtccctgc ccaccccgcc agcagctggt gttcctgaag acacataaat ccgggagcag 720  
ctctgtgctg agcctgcttc accgctatgg ggaccagcac gggtgcgct tcgcccctccc 780  
tgcccgcctac cagtttggct acccaagct cttccaggcc tctagggtaa aaggctaccg 840  
cccacagggt ggaggcacc agctcccctt ccacatctc tgtcaccaca tgaggttcaa 900  
cctgaaagag gtaactcagg tcatgccttc tgacagcttc tttttttcca ttgtccgaga 960  
cccagcggct ctggctcgtc ctgcctcttc ctactataaa tccacctcat cagccttccg 1020  
caagtcacca tctttggctg ccttccctggc caatcctcga ggcttctaca ggcctggggc 1080  
ccgtggggac cactacgctc gcaacttact atggtttgac ttggcctgc ccttccccc 1140  
agagaagagg gccaaagag ggaatttca tccccccaga gaccccaacc cccacagct 1200  
gcaggctctg ccttctgggt ctggccctcg agcccaaaacc ctcaatccca atgccctcat 1260  
ccatcctgtt tccactgtta ctgatcatcg cagccagata tcaagccctg cctctttcga 1320  
tttggggtct tcatccttca tccagtgggg tctggcctgg ctggactctg tctttgacct 1380  
ggtcatgggt gctggtactc tcatgagtc attggttctg ctggcagatg cctgtgctg 1440  
gggtctagat gacgtgggtg gcttcatgca caatgcccag gctggacata agcagggct 1500  
cagcactgtc agcaacagtg gactgactgc ggaggaccgg cagctgactg caccggccc 1560  
agcctggaac aacctggact gggtctctca tgtccacttc aaccgcagtc tctgggcacg 1620  
gatagagaaa tacggccagg gccggctgca gacagctgtg gccgagctcc gggctcgcg 1680  
agaggcccta gogaaacatt gtctggtagg gggtgaggct tctgaccca aatacatcac 1740  
tgatcgccgg ttccgcccct tccagtttg gtcagctaag gttttgggct atatacttcg 1800  
gagtggaatt agcccccagg accaagagga atgtgagcgc ctagctacce ctgagctcca 1860  
gtacaaggag aagctggatg tcaagcagtt ccccccctacc gtctcactgc cctcaagac 1920  
ttcaaggcca ctctcccat aaacatcaga ctacagattt aggtggaaga gcagccatgt 1980  
ttgaagggca catgtgatga gtggggggca gcaagatgcc attctgcat ctcccagaag 2040  
ggatgagttt ttgtccgat gcaagcccc tcttcgctgg gctcccagca gtgcttccct 2100  
cctccaccct ccactcattt tgttcttcc ccccaacttt ttttttttg aaacggagtc 2160  
ttgtctgtgc cccaggtg gagtgcagtg gcatgatctc ggctcactgc aacctctgcc 2220  
tcccagggtc aagcgattct cctgctcag cctccagagt agctaggatt acagatacgt 2280  
gccaccatac ccggctaatt tt 2302

<210> 36  
<211> 3341  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 3387823CB1

<400> 36  
gaaacgaaga gaacctctg tccctggcag aatctgcatg tacattttct gtctgtcctt 60  
gtctctcttc ttcctgtctg gccattgca gagagtattg gaagtttcca accattgggt 120  
gtactctatg ctcatcctac ctcccttctg gaaagacagt gtggcagcgc cctgtctgtc 180  
tgcctactac cctgactgtg ttggcatgag cccctcctgc accagcaca accgcgcgc 240

```

tgccactggc aatgccagcc ctgggaagct ggagcactcc aaggetgccc tctccgtgca 300
cgttccagcc atgaaccgat acttccagcc tttctaccag cccaatgagt gtggcaaacg 360
cctctgtgtg aggccggatg tgatggaact ggatgagctc tatgagtttc cagagtattc 420
ccgagacccc accatgtacc tggctttgag aaacctcacc ctgcgactgt ggtatactaa 480
ctgcaaagaa gctcttaact ctcagaaatg tattcctcac atcatogtcc ggggtctcgt 540
gcgtattcga tgcgttcagg aagtggagag aatactgtat tttatgacca gaaaaggctc 600
catcaacact ggagttctca gcgtgggagc cgaccagtat cttctcccta aggactacca 660
caataaatca gtcattcatta tcggggctgg tccagcagga tttagcagctg ctaggcaact 720
gcataacttt ggaattaagg tgaactgtcc ggaagccaaa gacagaattg gaggccgagt 780
ctgggatgat aaatctttta aaggcgtcac agtgggaaga ggagctcaga ttgtcaatgg 840
gtgtattaac aaccagtag cattaatgtg tgaacaactt ggcacagca tgcataaatt 900
tggagaaaga tgtgacttaa ttcaggaagg tggagaataa actgacccca ctattgacaa 960
gcgcattgat tttcatttta atgctctctt ggatgttgct tctgagtgga gaaaggataa 1020
gactcagctc caagatgtcc ctttaggaga aaagatagaa gaaatctaca aggcatttat 1080
taaggaaatc ggtatccaat tcagtgagct ggagggacag gtgcttcagt tccatctcag 1140
taacctggag taogcctgtg gcagcaacct tcaccaggta tctgctcgtc cgtgggacca 1200
caatgaattc tttgccaggt ttgctggtga ccacactctg ctaactcccg ggtactcggg 1260
gataattgaa aaactggcag aagggttga cattcaactc aaatctccag tgcagtgtat 1320
tgattattct ggagatgaag tgcaggttac cactacagat ggcacagggt attctgcaca 1380
aaaggtatta gtcactgtac cactggcttt actacagaaa ggtgccattc agtttaatcc 1440
accgttgtca gagaagaaga tgaaggctat caacagctta ggcgcaggca tcattgaaaa 1500
gattgccttg caatttccgt atagattttg ggacagtaaa gtacaagggg ctgacttttt 1560
tggtcacgtt cctcccagtg ccagcaagcg agggcttttt gccgtgttct atgacatgga 1620
tcccagaag aagcacagcg tgcgtatgtc tgtgattgcc ggggaggctg tgcacccgt 1680
gaggaccctg gacgacaaac aggtgctgca gcagtgcag gccacgctcc gggagctgtt 1740
caaggagcag gagggtccag atcccacaaa gtattttgtc actcgggtgga gcacagaccc 1800
atggatccag atggcataca gttttgtgaa gacaggtgga agtggggagg cctacgatat 1860
cattgtctgaa gacattcaag gaaccgtctt tttcgctggt gaggcaacaa acaggcattt 1920
cccacaaact gttacagggg catatttgag tggcgttcga gaagcaagca agattgcagc 1980
attttaagaa ttcgggtggc ccagctttct tctgtacccc agatggggaa atttgaatca 2040
catgttaaac ctcagtttta taagaggggg aaaaaaccgt ctctacatag taaaactgaa 2100
atgtttctaa ggcgatatga taatgcaaac ctatttcac actctaaaag cactgacctc 2160
aaaaaaacct ataagcactt agatttaatt gcattttcca taggttcaac tactgtgaa 2220
agtctggatt tcagaataaa gcagaatgta agtttcagtt gaggccatgg atttgattgt 2280
tccatggctg gaagttccct ttagatttca cattttatat ggtgatcaa ttttcataca 2340
ttgagaaacc aagtcaatca agcaggaatc atttaaaaac cagataaagc catgtttttc 2400
ttctgtgaca atttatcagt atctttacca atgagcetta atttttatat aggtccaata 2460
ttgagctttt acttaaaatt tagatagaac ttttttttgg atacagcaca aactccagtt 2520
gacagtaaaa tgaagcttct aggtattttg tattgtacat atttccctct actgggtgtt 2580
caaaagaaat ttaaatcaa gtaccttttg tgataaaatg ttttagattt gtgcacccat 2640
tggcaaaaca ggaaagtttc cagataggta ttgtatcatt gagaatgcag cacagatagt 2700
gtgggcttca cactatagac acagaatata gctttttctt aaagccaaat ttgggtgata 2760
ggacacttta aatatacctta attttggcaa ccactagcaa aaaaaacttg tcagaataat 2820
ttaaccaagc cctctccac ttcttttatt taaaagcact gattcaattg ctaggatat 2880
ttttgcagat ttttctttac agtattocat aggcaggctc actggaaaac tgcagaaaaa 2940
tgtgagctct cctggtaaat agtatacatt ttataageta tatttttaag gcctaagaa 3000
atggcgagta tttactttta tctttttttt aaaaacactc atgacagaaa acagtctaat 3060
aatatctcat tctaaaataa aacactgggt gcagggtctt caggatgcct attttgccag 3120
aaacttcagt atacaggtta gaaatatgct tttgtttttg aacataatat actggttttc 3180
tttaaaagaag ggactaaata tgacttttaa gagacttcaa atattgagta ttttaaaaat 3240
ttaaaagtag gtcagtttat aacgagtaaa tacctaacac ccaagaatgt gcagtgaacc 3300
tcaggcgggg atccttagtt ctaacggccg ccccgtaggc g 3341

```

&lt;210&gt; 37

&lt;211&gt; 777

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 55142051CB1

&lt;400&gt; 37

```

atgtcccttg ccattgcatt ggcttctctg ccactggtgg taacattgct ggtgcggtac 60
cggcactact tccgattgct ggtgcgcacg gtcttctctg gaagcctccg agactgcctg 120
tcagggtctg ggatcgagga gggggccttc agctacgtgc tcacccatgc cctgcccgtg 180
gaccctgggc acatcctcac caccctggac cactggagca gccgtgcga gtacttgagc 240
cacatggggc ctgtcaaagg tcagatcctg atgcggctgg tggaggagaa ggcccctgct 300
tgtgtgctgg aattgggaac ctactgtgga tactctaccc tgcttattgc ccgagccctg 360
ccccctgggg gtgccttctt tactgtggag cgggacccac gcacggcagc agtggctgaa 420
aaactcatcc gcctggcggg ctttgatgag cacatggtgg agctcatcgt gggcagctca 480
gaggacgtga tcccgctgct acgcacccag tatcagctga gtcgggcaga cctgggtgctc 540
ctggcacacc ggccacgatg ttacctgagg gacctgcagc tgctggaggc ccatgcccta 600
ctgccagcag gtgccaccgt gctggctgac catgtgctct tccctgggtg acccgccttc 660
ttgcagtatg ctaagagctg tggcgcctac cgctgcggcc tccaccacac tggccttcca 720
gacttccctg ccatcaagga tggaatagct cagctcacct atgctggacc aggtgta 777

```

&lt;210&gt; 38

&lt;211&gt; 3600

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7395274CB1

&lt;400&gt; 38

```

gagagcagag caggacactg gcgcgcgggg tcaggcagct gcgtgcgctg ctccctccagg 60
oagcaagggg aacccgaggc cgcggcgccc cggaccatgt cgtctccggg gccgtgcag 120
ccgcggcgcc aggaccggcc ctggcccggg cgtctccctg gtgcgcctct ggggctgctg 180
cggtgggacc ccagcggggg cgcgtgctg ctatgcggcc tcgtagcgtg gctgggctgg 240
agctggctgc ggaggcgccg ggcgcggggc atcccgcccg ggcccacgcc ctggcctctg 300
gtgggcaact tcggtcacgt gctgctgcct cccttctctc ggcggcggag ctggctgagc 360
agcaggacca ggcgcgcagg gattgatccc tcggctatag gcccgagggt gctcctggct 420
cacctagccc gcgtgtacgg cagcatcttc agcttcttta tcggccacta cctgggtggtg 480
gtcctcagcg acttccacag cgtgcgcgag gcgctggtgc agcaggccga ggtcttcagc 540
gaccgcccgc ggggtgcgct catctccatc gtgaccaagg agaagggggg tgtgtttgca 600
cattattggtc ccgtctggag acaacaaagg aagttctctc attcaactct tcgtcatttt 660
gggttgggaa aacttagctt ggagcccaag attattgagg agttcaata tgtgaaagca 720
gaaatgcaaa agcacggaga agacccttcc tgccctttct ccatcatcag caatgccgtc 780
totaacatca tttgctcctt gtgctttggc cagcgctttg attacactaa tagtgagttc 840
aagaaaatgc ttggttttat gtcacgaggc ctagaatct gtctgaacag tcaagtcctc 900
ctggtcaaca tatgccttgg cttttattac ctcccttttg gaccatttaa ggaattaaga 960
caaattgaaa aggatataac cagtttctct aaaaaaatca tcaaagacca tcaagagtct 1020
ctggatagag agaaccctca ggacttcata gacatgtacc ttctccacat ggaagaggag 1080
aggaaaaata atagtaacag cagttttgat gaagagtact tattttatat cattggggat 1140
ctctttattg ctgggactga taccacaact aactctttgc tctggtgcct gctgtatatg 1200
tcgctgaacc ccgatgtaca agaaaagggt catgaagaaa ttgaaagagt cattggcgcc 1260
aaccgagctc ctccctcac agacaaggcc cagatgcctt acacagaagc caccatcatg 1320
gaagtgcaga ggctaactgt ggtggtgccc ctgcccattc ctcatatgac ctccagagaa 1380
acagtgtctc aagggtatata catttctaaa ggcacattga tcttacccaa cctgtggtca 1440
gtacatagag acccagccat ttgggagaaa ccggaggatt tctaccctaa tcgattttctg 1500
gatgaccaag gacaactaat taaaaaagaa acctttatct cttttgggat agggaagcgg 1560
gtgtgtatgg gagaacaact ggcaagatg gaattattcc taatgtttgt gaggcctaag 1620
cagagtctcg catttgcttt acctgaggat tctaagaagc ccctctgac tggagagatt 1680
ggtctaactt tagcccccaca tccatttaata ataactatct caaggagatg aagagcatct 1740
ccaagaagag atggtaaaaa gatataataa tacatatcct tctaagcaga ttcttctac 1800
tgcaaaggac agtgaatcca gcaactcagt ggatccaagc tgggctcaga ggtcggaagg 1860
agggtagagc acactgggag gtttcatctt ggaggattcc tcagcaggat acttcagcca 1920
tttttagtaat gcaggtctgt gatttggggg atagaaaaca aagtacctat gaaacgggat 1980
atctgatttt tacttgaggt ggcttccacc gatgggcaaa tcttctcatt tcttagtgcc 2040
tcagacatcc catatgtaaa atgagagtaa taaaacttgg cttctctcta cctctcagca 2100
ctaagtatgg tcaaatgcct tacatctttt ctgatatctc taaaatgctg ttaagttctg 2160
gagaagaact tcaggagaag aagatctatc agctggcttt taaagacctc tgacaacatg 2220

```

aaagtgggtg tcagcctgga atgctttgtc agagatgggt gtggatttag gttatactgg 2280  
gggagaaact ttctcagcac agattctatg ccagcttctt tgggcttggt ctgtcactat 2340  
ctttttgttt atgatttttag tttttacttt ttgtagatgt gggatgaagt ggactctgtc 2400  
gtgtatatatt aggaaaaaag aaattataat tttaaaaaat cccttgtagg attattatct 2460  
aaatttataat gtctaacttc tactacaact acaggaacag tgagccttgc tacttcttta 2520  
gtagcttctt ggcagaattc ctttctactg agttatttgc aaagatgcag ctctaccttt 2580  
ttacttaagg cctgaatggg gagcatgggg attttgatac tgggactcat caggaaagga 2640  
ttctgctttc aaactatact gaacattcct gtctagcgt ccctgccacc aggcccaatg 2700  
catctgatec ttgaatatac tctcaaagaa ttcactctct ttttattaag agaactaaat 2760  
tgtttctaaa tgtagatggg ccctctggaa aagcagtttt cagcaggggt ggtaaccct 2820  
tcagagggag tttggaaatg tgtgggtatg attcttggtt atcataatga tgggggtgct 2880  
actggccttc tgcctccatg ggaccaggat gctaaatgct aaggtagtcc tatacagtga 2940  
agaattgtcc tgctcaagat gccaggattt cccccagtga gaacatgctc taaggatga 3000  
ccacccttt cttttattct cccacagtgc tccatgtaca gaagtaagca tagcagtc 3060  
atgagcaacc acattcctga acctttctc atgctggctc tacacttaat cctttacttg 3120  
tatgtttctg taattcttac ataaattcta ttaagagggt ggcatactgt agtggatgaa 3180  
gctgaggctt atagtaggta aggcacaaag ttaaaaagta acatcactgg gtttcaaacc 3240  
tactggctc tgtgactaaa gaacacttcc agaaccactt cttgattctg ccaccacttg 3300  
atcccataac aggctacccc ttggcctcat gctggagttg tgtgtgtctg tcttcacccc 3360  
aggctgagct ccttgagggt aggatgttgt gctgtttgcc tccctcacag tgcttgggtc 3420  
ttagtggatg cccagttgtc ttgtgaatga cttttaagaa gtgtacttaa gagaaaaatc 3480  
ctaccttatt tgaataatta caagtcatgt ttttgttgct taaaggatg aaatcagtg 3540  
atattatttg ttaatgtcca ttaaagccag tttttaaaaa aaaaaaaaaa aaaaaaaaaa 3600